03-17-06

AF/1647

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In application of:	Examiner: Romeo, David
Ashkenazi, et al.	Art Unit: 1647
Application Serial No. 09/904,838	Confirmation No: 5331
Filed: July 13, 2001	Attorney's Docket No. 39780-1618 P2C17
For: **SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME**	**Customer No. 35489**

EXPRESS MAIL LABEL NO. : EL 976 551 904 US
DATE MAILED: March 15, 2005

ON APPEAL TO THE BOARD OF PATENT APPEALS AND INTERFERENCES

APPELLANTS' BRIEF

MAIL STOP APPEAL BRIEF - PATENTS
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 11413-1450

Dear Sir:

This Appeal Brief, filed in connection with the above captioned patent application, is responsive to the Final Office Action mailed on June 15, 2005. A Notice of Appeal was filed herein on September 15, 2005. A request for a four month extension of time is requested herewith. Appellants hereby appeal to the Board of Patent Appeals and Interferences from the final rejection in this case.

The following constitutes the Appellants' Brief on Appeal.

I. REAL PARTY IN INTEREST

The real party in interest is Genentech, Inc., South San Francisco, California, by an assignment of the parent application, U.S. Patent Application Serial No. 09/665,350 recorded July 9, 2001, at Reel 011964 and Frame 0181. The present application is a continuation of U.S. Serial No. 09/665,350.

II. RELATED APPEALS AND INTERFERENCES

The claims pending in the current application are directed to a polypeptide referred to herein as "PRO317". There exists one related patent application, U.S. Serial No. 09/906,760, filed November 19, 2001 (containing claims directed to nucleic acids encoding PRO317 polypeptides).

III. STATUS OF CLAIMS

Claims 44-46 and 49-51 are in this application.

Claims 1-44 and 47-48 have been canceled.

Claims 44-46 and 49-51 stand rejected and Appellants appeal the rejection of these claims.

A copy of the rejected claims in the present Appeal is provided as Appendix A.

IV. STATUS OF AMENDMENTS

There were no amendments submitted after the final rejection mailed December 23, 2004. All previous amendments have been entered.

V. SUMMARY OF CLAIMED SUBJECT MATTER

The invention claimed in the present application is related to an isolated polypeptide comprising the amino acid sequence of the polypeptide of SEQ ID NO: 114, referred to in the present application as "PRO317." The PRO317 gene was shown for the first time in the present application to be significantly amplified in human lung cancers as compared to normal, non-cancerous human tissue controls (Example 92). This feature is specifically recited in claim 124, and carried by all claims dependent from claim 44. In addition, the invention also claims the amino acid sequence of the polypeptide of SEQ ID NO: 114, lacking its associated signal-

peptide; or the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 203025 (Claim 44-46 and 49). The invention is further directed to a chimeric polypeptide comprising one of the above polypeptides fused to a heterologous polypeptide (Claim 50), and to a chimeric polypeptide wherein the heterologous polypeptide is an epitope tag or an Fc region of an immunoglobulin (Claim 51). The preparation of chimeric PRO polypeptides (claims 50 and 51), including those wherein the heterologous polypeptide is an epitope tag or an Fc region of an immunoglobulin, is set forth in the specification at page 74, lines 23 to page 75, line 5. Examples 53-56, pages 192-199, describe the expression of PRO polypeptides in various host cells, including *E. coli*, mammalian cells, yeast and Baculovirus-infected insect cells.

The amino acid sequence of the "PRO317" polypeptide and the nucleic acid sequence encoding this polypeptide (referred to in the present application as "DNA33461-1199") are shown in the present specification as SEQ ID NOs: 114 and 113, respectively, and in Figures 42 and 41, described on page 60, lines 30-33. The full-length PRO317 polypeptide having the amino acid sequence of SEQ ID NO:114 is described in the specification at, for example, on page 15, page 41 and pages 103-104, page 133, line 16 to page 135, line 18 and the isolation of cDNA clones encoding PRO317 of SEQ ID NO:114 is described in Example 18, page 162 of the specification. The specification discloses that various portions of the PRO317 polypeptide possess significant sequence similarity to EBAF which is expressed in the late secretory phase of endometrial bleeding and belongs to the TGF-β superfamily of proteins and PRO317 polypeptides and compositions thereof maybe useful in diagnosing and treating abnormal bleeding conditions in the endometrium for instance (see, for example, page 16-17 and page 133, line 16 to page 135, line 18 and Example 18).

Finally, Example 92, in the specification at page 222, line 26, to page 235, line 3, sets forth a 'Gene Amplification assay' which shows that the PRO317 gene is amplified in the genome of certain human lung cancers (see Table 9B, page 231). The profiles of various primary lung and colon tumors used for screening the PRO polypeptide compounds of the invention in the gene amplification assay are summarized on Table 8, page 227 of the specification.

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

1. Whether International Application PCT/US98/18824, filed September 10, 1998 satisfies the utility/ enablement requirement under 35 U.S.C. §101/112, first paragraph and whether the instant Claims 44-46 and 49-51 are entitled to the priority of Application PCT/US98/18824.

2. Whether Claims 44-46 and 49-51 are anticipated under 35 U.S.C. §102(a) by Ruben *et al.*, WO 99/09198 (February 1999).

VII. ARGUMENTS

Summary of the Arguments

Issue 1: Priority

The sole basis for the Examiner's rejection of the priority date of earlier filed Application, PCT/US98/18824 (September 10, 1998) is because the subject matter presented in this earlier specification is allegedly insufficient under 35 U.S.C. § 112, first paragraph.

Since the 'how to use' prong of the enablement requirement under 35 U.S.C. §112, first paragraph incorporates, as a matter of law, a requirement that the specification disclose a practical utility for the claimed invention the utility requirements under 35 U.S.C. §101 are discussed.

Appellants have previously submitted that patentable utility for the PRO317 polypeptides is based upon the gene amplification data for the gene encoding the PRO317 polypeptide. The specification discloses that the gene encoding PRO317 showed significant amplification, ranging from **2.028 to 6.774**-fold in 8 different lung primary tumors and from **2.06 to 6.73**-fold in 6 different colon primary tumors. Appellants submit that the PRO317 polypeptide is useful as a marker for the diagnosis of lung or colon cancer, and for monitoring cancer development and/or for measuring the efficacy of cancer therapy.

While the Examiner agrees and says that "(t)he examiner is not arguing that a correlation between PRO317 gene amplification and (PRO317) polypeptide expression does not exist" "(Page 2, last few lines through page 3 of the Final Office Action mailed June 15, 2005), he adds, based on Pennica *et al.*, Haynes *et al.* and Hancock *et al.* that "the analysis of protein products is essential because protein expression levels are not predictable from the mRNA expression levels" (Page 5, lines 11-12 of the Final Office Action mailed June 15, 2005).

Appellants submit that the teachings of Pennica *et al.*, Haynes *et al.* and Hancock *et al.* do not conclusively establish a *prima facie* case for lack of utility because the references are either not contrary to the Appellants' arguments or actually lend support to the Appellants' position, as explained in detail below.

In addition, Appellants have submitted ample evidence to show that, in general, if a gene is amplified in cancer, it is more likely than not that the encoded protein will be expressed at an elevated level. First, the articles by Ornoft *et al.*, Hyman *et al.*, and Pollack *et al.* (made of record in Appellants' Response filed June 28, 2004) collectively teach that in general, gene amplification increases mRNA expression. Second, the Declaration of Dr. Paul Polakis (made of record in Appellants' Response filed June 28, 2004), principal investigator of the Tumor Antigen Project of Genentech, Inc., the assignee of the present application, shows that, in general, there is a correlation between mRNA levels and polypeptide levels . Appellants further note that the sale of gene expression chips to measure mRNA levels is a highly successful business, with a company such as Affymetrix recording 168.3 million dollars in sales of their GeneChip arrays in 2004. Clearly, the research community believes that the information obtained from these chips is useful (i.e., that it is more likely than not informative of the protein level).

Taken together, although there are some examples in the scientific art that do not fit within the central dogma of molecular biology that there is a correlation between DNA, mRNA, and polypeptide levels, these instances are exceptions rather than the rule. In the majority of amplified genes, as exemplified by Ornoft *et al.*, Hyman *et al.*, Pollack *et al.*, the Polakis Declaration and the widespread use of array chips, the teachings in the art overwhelmingly show that gene amplification influences gene expression at the mRNA and protein levels. Therefore, one of skill in the art would reasonably expect in this instance, based on the amplification data for the PRO317 gene, that the PRO317 polypeptide is concomitantly overexpressed. Thus, the claimed PRO317 polypeptides also have utility in the diagnosis of lung or colon cancer.

Appellants further submit that even if there is no correlation between gene amplification and increased mRNA/protein expression, (which Appellants expressly do not concede), a polypeptide encoded by a gene that is amplified in cancer would still have a specific, substantial, and credible utility. Appellants submit that, as evidenced by the Ashkenazi Declaration and the teachings of Hanna and Mornin (both made of record in Appellants' Response filed June 28, 2004), simultaneous testing of gene amplification and gene product over-expression enables

more accurate tumor classification, even if the gene-product, the protein, is not over-expressed. This leads to better determination of a suitable therapy for the tumor, as demonstrated by a real-world example of the breast cancer marker HER-2/neu. Accordingly, Appellants submit that when the proper legal standard is applied, one should reach the conclusion that the present application discloses at least one patentable utility for the claimed PRO317 polypeptides.

Therefore, Appellants submit that the earlier filed Application PCT/US98/18824 (September 10, 1998) satisfies 35 U.S.C. § 112, first paragraph and the present application is entitled to the earlier filing date of **September 10, 1998**.

Issue 2: Anticipation by Ruben *et al.*, WO99/09198

The instant application claims PRO317 polypeptides. As discussed above, the present application is entitled to the earlier filing date of **September 10, 1998** and therefore, Ruben *et al.*, WO99/09198, dated February 1999, is not prior art for the instant application. Thus the instant claims are not anticipated by Ruben *et al.*

These arguments are all discussed in further detail below under the appropriate headings.

Response to Rejections

ISSUE 1. International Application PCT/US98/18824 Satisfies the Utility Requirement of 35 U.S.C. § 101 / § 112, First Paragraph based on the results of the gene amplification assay

The sole basis for the Examiner's rejection of Claims 44-46 and 49-51 under this section is that the data presented in the earlier filed Application, PCT/US98/18824 (September 10, 1998) is allegedly insufficient under 35 U.S.C. § 112, first paragraph. Since the 'how to use' prong of the enablement requirement under 35 U.S.C. §112, first paragraph incorporates, as a matter of law, a requirement that the specification disclose a practical utility for the claimed invention the utility requirements under 35 U.S.C. §101 are discussed.

Appellants strongly disagree and respectfully traverse the rejection.

A. The Legal Standard For Utility Under 35 U.S.C. § 101

According to 35 U.S.C. § 101:

> Whoever invents or discovers any new and *useful* process, machine, manufacture, or composition of matter, or any new and *useful* improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title. (Emphasis added.)

In interpreting the utility requirement, in *Brenner v. Manson*[1] the Supreme Court held that the *quid pro quo* contemplated by the U.S. Constitution between the public interest and the interest of the inventors required that a patent applicant disclose a "substantial utility" for his or her invention, i.e. a utility "where specific benefit exists in currently available form."[2] The Court concluded that "a patent is not a hunting license. It is not a reward for the search, but compensation for its successful conclusion. A patent system must be related to the world of commerce rather than the realm of philosophy."[3]

Later, in *Nelson v. Bowler*[4] the C.C.P.A. acknowledged that tests evidencing pharmacological activity of a compound may establish practical utility, even though they may not establish a specific therapeutic use. The court held that "since it is crucial to provide researchers with an incentive to disclose pharmaceutical activities in as many compounds as possible, we conclude adequate proof of any such activity constitutes a showing of practical utility."[5]

In *Cross v. Iizuka*[6] the C.A.F.C. reaffirmed *Nelson*, and added that *in vitro* results might be sufficient to support practical utility, explaining that "*in vitro* testing, in general, is relatively less complex, less time consuming, and less expensive than *in vivo* testing. Moreover, *in vitro* results with the particular pharmacological activity are generally predictive of *in vivo* test results, i.e. there is a reasonable correlation there between."[7] The court perceived "No insurmountable

---

[1] *Brenner v. Manson,* 383 U.S. 519, 148 U.S.P.Q. (BNA) 689 (1966).

[2] *Id.* at 534, 148 U.S.P.Q. (BNA) at 695.

[3] *Id.* at 536, 148 U.S.P.Q. (BNA) at 696.

[4] *Nelson v. Bowler*, 626 F.2d 853, 206 U.S.P.Q. (BNA) 881 (C.C.P.A. 1980).

[5] *Id.* at 856, 206 U.S.P.Q. (BNA) at 883.

[6] *Cross v. Iizuka*, 753 F.2d 1047, 224 U.S.P.Q. (BNA) 739 (Fed. Cir. 1985).

[7] *Id.* at 1050, 224 U.S.P.Q. (BNA) at 747.

difficulty" in finding that, under appropriate circumstances, "in vitro testing, may establish a practical utility."[8]

The case law has also clearly established that Appellants' statements of utility are usually sufficient, unless such statement of utility is unbelievable on its face.[9] The PTO has the initial burden to prove that Appellants' claims of usefulness are not believable on their face.[10] In general, an Applicant's assertion of utility creates a presumption of utility that will be sufficient to satisfy the utility requirement of 35 U.S.C. §101, "unless there is a reason for one skilled in the art to question the objective truth of the statement of utility or its scope."[11, 12]

Compliance with 35 U.S.C. §101 is a question of fact.[13] The evidentiary standard to be used throughout *ex parte* examination in setting forth a rejection is a preponderance of the totality of the evidence under consideration.[14] Thus, to overcome the presumption of truth that an assertion of utility by the applicant enjoys, the Examiner must establish that it is more likely than not that one of ordinary skill in the art would doubt the truth of the statement of utility. Only after the Examiner made a proper *prima facie* showing of lack of utility, does the burden of rebuttal shift to the applicant. The issue will then be decided on the totality of evidence.

---

[8] *Id.*

[9] *In re Gazave*, 379 F.2d 973, 154 U.S.P.Q. (BNA) 92 (C.C.P.A. 1967).

[10] *Ibid.*

[11] *In re Langer*, 503 F.2d 1380,1391, 183 U.S.P.Q. (BNA) 288, 297 (C.C.P.A. 1974).

[12] *See also In re Jolles*, 628 F.2d 1322, 206 USPQ 885 (C.C.P.A. 1980); *In re Irons*, 340 F.2d 974, 144 USPQ 351 (1965); *In re Sichert*, 566 F.2d 1154, 1159, 196 USPQ 209, 212-13 (C.C.P.A. 1977).

[13] *Raytheon v. Roper*, 724 F.2d 951, 956, 220 U.S.P.Q. (BNA) 592, 596 (Fed. Cir. 1983) cert. denied, 469 US 835 (1984).

[14] *In re Oetiker*, 977 F.2d 1443, 1445, 24 U.S.P.Q.2d (BNA) 1443, 1444 (Fed. Cir. 1992).

The well established case law is clearly reflected in the Utility Examination Guidelines ("Utility Guidelines")[15], which acknowledge that an invention complies with the utility requirement of 35 U.S.C. §101, if it has at least one asserted "specific, substantial, and credible utility" or a "well-established utility." Under the Utility Guidelines, a utility is "specific" when it is particular to the subject matter claimed. For example, it is generally not enough to state that a nucleic acid is useful as a diagnostic without also identifying the conditions that are to be diagnosed.

In explaining the "substantial utility" standard, M.P.E.P. §2107.01 cautions, however, that Office personnel must be careful not to interpret the phrase "immediate benefit to the public" or similar formulations used in certain court decisions to mean that products or services based on the claimed invention must be "currently available" to the public in order to satisfy the utility requirement. "Rather, any reasonable use that an applicant has identified for the invention that can be viewed as providing a public benefit should be accepted as sufficient, at least with regard to defining a 'substantial' utility."[16] Indeed, the Guidelines for Examination of Applications for Compliance With the Utility Requirement,[17] gives the following instruction to patent examiners: "If the applicant has asserted that the claimed invention is useful for any particular practical purpose . . . and the assertion would be considered credible by a person of ordinary skill in the art, do not impose a rejection based on lack of utility."

B. <u>Proper Application of the Legal Standard</u>

Appellants respectfully submit that the data presented in Example 92 starting on page 222 of the priority application and the cumulative evidence of record, which underlies the current dispute, indeed support a "specific, substantial and credible" asserted utility for the presently claimed invention.

Patentable utility for the PRO317 polypeptides is based upon the gene amplification data for the gene encoding the PRO317 polypeptide. Example 92 describes the results obtained using

---

[15] 66 Fed. Reg. 1092 (2001).

[16] M.P.E.P. §2107.01.

[17] M.P.E.P. §2107 II (B)(1).

a very well-known and routinely employed polymerase chain reaction (PCR)-based assay, the TaqMan™ PCR assay, also referred to herein as the gene amplification assay. This assay allows one to quantitatively measure the level of gene amplification in a given sample, say, a tumor extract, or a cell line. It was well known in the art at the time the invention was made that gene amplification is an essential mechanism for oncogene activation. Appellants isolated genomic DNA from a variety of primary cancers and cancer cell lines that are listed in Table 9 (pages 222 onwards of the specification), including primary lung and colon cancers of the type and stage indicated in Table 8 (page 227). The tumor samples were tested in triplicates with Taqman™ primers and with internal controls, beta-actin and GADPH in order to quantitatively compare DNA levels between samples (page 229). As a negative control, DNA was isolated from the cells of ten normal healthy individuals, which was pooled and used as a control (page 222, lines 28-29). The results of TaqMan™ PCR are reported in ΔCt units, as explained in the passage on page 222, lines 37-39. One unit corresponds to one PCR cycle or approximately a 2-fold amplification, relative to control, two units correspond to 4-fold, 3 units to 8-fold amplification and so on . Using this PCR-based assay, Appellants showed that the gene encoding for PRO317 was amplified, that is, it showed approximately 1.02-2.76 ΔCt units for lung tumors and 1.04-2.75 ΔCt units for colon tumors, which corresponds to $2^{1.02}$-$2^{2.76}$- fold amplification in lung or to $2^{1.04}$-$2^{2.75}$- fold amplification in colon tumors; that is **2.028 to 6.774**-fold in 8 different lung primary tumors and from **2.06 to 6.73**-fold in 6 different colon primary tumors, which would be considered significant and credible by one skilled in the art. Therefore, the PRO317 gene and the PRO317 polypeptide are important diagnostic markers to identify such malignant lung or colon cancers.

The Examiner says that "(t)he examiner is not arguing that a correlation between PRO317 gene amplification and (PRO317) polypeptide expression does not exist" "(Page 2, last few lines through page 3 of the Final Office Action mailed June 15, 2005). Thus, the Examiner seems to agree that a correlation exists between DNA and protein expression levels in general. But the Examiner points out that,

> "the present specification fails to disclose what that correlation is or the significance of any such correlation. The specification fails to disclose enough information about the invention to make its usefulness immediately apparent to those familiar with the technological field of the invention. (page 2, last line to page 3, line 3)"

The Examiner adds that "Pennica is evidence that **not all gene amplifications** are associated with overexpression of the corresponding gene product and that the skilled artisan would not have appreciated that PRO317 gene amplification without more (emphasis added).

Appellants strongly disagree. Appellants submit that the Examiner applied an improper legal standard when making this rejection. The evidentiary standard to be used throughout *ex parte* examination of a patent application is a preponderance of the totality of the evidence under consideration. Thus, to overcome the presumption of truth that an assertion of utility by the applicant enjoys, the Examiner must establish that it is more likely than not that one of ordinary skill in the art would doubt the truth of the statement of utility. Only after the Examiner has made a proper *prima facie* showing of lack of utility, does the burden of rebuttal shift to the applicant. Accordingly, it is not a legal requirement to establish a necessary correlation between an increase in the copy number of the DNA and protein expression levels nor is it imperative to find evidence that DNA amplification is "necessarily" or "always" associated with overexpression of the gene product. Appellants respectfully submit that when the proper evidentiary standard is applied, a correlation must be acknowledged.

First of all, the teachings of Pennica *et al.* are specific to *WISP* genes, a specific class of closely related molecules. Pennica *et al.* showed that there was good correlation between DNA and mRNA expression levels for the *WISP-1* gene but not for *WISP-2 and WISP-3* genes. But, the fact that in the case of closely related molecules, there seemed to be no correlation between gene amplification and the level of mRNA/protein expression does not establish that it is more likely than not, in general, that such correlation does not exist. As discussed above, the standard is not absolute certainty. Pennica *et al.* has no teaching whatsoever about the correlation of gene amplification and protein expression for genes in general. Indeed, the working hypothesis among those skilled in the art is that, if a gene is amplified in cancer, the encoded protein is likely to be expressed at an elevated level. In fact, as noted even in Pennica *et al.*, "[a]n analysis of *WISP*-1 gene amplification and expression in human colon tumors *showed a correlation between DNA amplification and over-expression* . . . " (Pennica *et al.*, page 14722, left column, first full paragraph, emphasis added). Accordingly, Appellants respectfully submit that Pennica *et al.* teaches nothing conclusive regarding the absence of correlation between gene amplification and over-expression of mRNA or polypeptides in most genes, in general.

The Examiner adds that "Haynes *et al.* and Hancock *et al.* provide evidence that Dr. Polakis' asserted dogma is not absolutely true" (Page 4, line 1 of the Final Office Action mailed June 15, 2005).

First, regarding the non-acceptance of the Polakis and other declarations by the Examiner, Appellants respectfully draw the Examiner's attention to case law that clearly establishes that in considering affidavit evidence, the Examiner must consider all of the evidence of record anew (*In re Rinehart,* 531 F.2d 1084, 189 USPQ 143 (C.C.P.A. 1976) and *In re Piasecki,* 745 F.2d. 1015, 226 USPQ 881 (Fed. Cir. 1985)). "After evidence or argument is submitted by the applicant in response, patentability is determined on the totality of the record, by a preponderance of the evidence with due consideration to persuasiveness of argument" (*In re Alton,* 37 USPQ2d 1578 (Fed. Cir 1966) at 1584 quoting *In re Oetiker,* 977 F.2d 1443, 1445, 24 USPQ2d 1443, 1444 (Fed. Cir. 1992)). Furthermore, the Federal Court of Appeals held in *In re Alton*, "We are aware of no reason why opinion evidence relating to a fact issue should not be considered by an examiner" (*In re Alton, supra.*). Appellants further draw the Examiner's attention to the Utility Examination Guidelines (Part IIB, 66 Fed. Reg. 1098 (2001)) which states,

> "Office personnel must accept an opinion from a qualified expert that is based upon relevant facts whose accuracy is not being questioned; it is improper to disregard the opinion solely because of a disagreement over the significance or meaning of the facts offered."

The statement in question from the Polakis Declaration that "it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell" is based on his own experimental findings, which is clearly set forth in the Declaration. Further, Appellants add that the teachings of Ashkenazi were supported by the Her-2/neu gene example in Hanna and Mornin. Accordingly, the fact-based conclusions of Dr. Polakis and Dr. Ashkenazi would be considered reasonable and accurate by one skilled in the art.

Regarding Haynes, the Examiner quotes Haynes as follows:

> "These results suggest that even for a population of genes predicted to be relatively homogenous....., the protein levels **cannot be accurately predicted** from the level of the

corresponding mRNA transcript" (page 1863, left column, full paragraph 1) (Final Office action, page 4, last paragraph).

Therefore the Examiner concludes that protein levels cannot be **accurately predicted** from the level of the corresponding mRNA transcript.

First of all, Appellants submit that it is not a legal requirement to accurately predicted from the level of the corresponding mRNA transcript or to establish a necessary or "strong" correlation between an increase in the DNA/ mRNA copy number and protein expression levels for an assertion of utility, nor is it imperative to find evidence that DNA amplifications are always associated with overexpression of the gene product. Instead, the question is whether it is more likely than not that a person of ordinary skill in the art would recognize a positive correlation.

Contrary to the Examiner's reading, Haynes *et al.* teaches that "there was a *general trend but no strong correlation* between protein [expression] and transcript levels" (Emphasis added). For example, in Figure 1, there is a positive correlation between mRNA and protein levels amongst most of the 80 yeast proteins studied. In fact, very few data points deviated or scattered away from the expected normal and no data points showed a negative correlation between mRNA and protein levels (*i.e.,* an increase in mRNA resulted in a decrease in protein levels). Haynes *et al.* notes that their analysis focused on the 80 most abundant proteins in the yeast lysate (page 1867). Haynes *et al.* states "since many important regulatory proteins are present only at low abundance, these would not be amenable to analysis" (page 1867). Haynes *et al.* compared the protein expression levels of these naturally abundant proteins to mRNA expression levels from published SAGE frequency tables. (page 1863). Thus, contrary to the Examiner's position, the Haynes data actually supports Polakis' statement that, in general, a positive correlation exists between mRNA and protein levels (even though the correlation may not be linear and hence, the data cannot be used to accurately predict protein levels or amounts). The Haynes data in fact, meets the "more likely than not" utility standard since it studied **80 proteins** and showed "a general positive trend" or increase in protein levels for most of the 80 proteins with corresponding mRNA increases.

Therefore, when the proper legal standard is used, a *prima facie* case of lack of utility has not been met based on the cited references Haynes *et al.* Indeed, the working hypothesis among

those skilled in the art is that, if a gene is amplified in cancer, the encoded protein is likely to be expressed at an elevated level.

Further, Appellants submit that the Hancock reference cited by the Examiner does not provide evidence that Dr. Polakis' statements are not absolutely true. Hancock discusses the need for high-quality biomarkers in the genomics and proteomics era and the need for a "consensus-building process" and "consolidation of different lists of biomarkers". While the editorial indicates that the markers generated by proteomics are not always consistent with markers identified by expression profiling, which possibly reflects methodological differences between expression and proteomic studies, the statements in the editorial by no means provide evidence that Dr. Polakis' statements are not absolutely true. In fact, the statements in the editorial indicate the importance for proteomics (and protein markers generated thereof) in the third paragraph: "I think many people in the proteomics community would agree that federal granting agencies *should be enticed* to continue investments in basic proteomics technology." Thus in fact, Hancock provides evidence that **biomarkers like PRO317 are useful**, and in fact desirable, provided there is a push towards a consolidated list of biomarkers (which is outside the scope of the utility requirement). Thus, Appellants respectfully point out that the Hancock reference in fact supports utility for protein markers despite seeming discrepancies between expression and proteomic studies.

On the contrary, Appellants submit that gene amplification assay in the specification further discloses that, "(a)mplification is associated with overexpression of the gene product, indicating that the polypeptides are useful targets for therapeutic intervention in certain cancers such as colon, lung, breast and other cancers and diagnostic determination of the presence of those cancers" (emphasis added). Besides, Appellants have submitted ample evidence (discussed below) to show that, in general, if a gene is amplified in cancer, it is "more likely than not" likely that the encoded protein will also be expressed at an elevated level.

For support, Appellants presented the articles by Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.* (made of record in Appellants' Response filed June 28, 2004), who collectively teach that in general, for most genes, DNA amplification increases mRNA expression. The results presented by Orntoft *et al.*, Hyman *et al.*, and Pollack *et al.* are based upon wide ranging analyses of a large number of tumor associated genes. Orntoft *et al.* studied transcript levels of 5600 genes in malignant bladder cancers, many of which were linked to the gain or loss of

chromosomal material, and found that in general (18 of 23 cases) chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts. Hyman *et al.* compared DNA copy numbers and mRNA expression of over 12,000 genes in breast cancer tumors and cell lines, and found that there was evidence of a prominent global influence of copy number changes on gene expression levels. In Pollack *et al.*, the authors profiled DNA copy number alteration across 6,691 mapped human genes in 44 predominantly advanced primary breast tumors and 10 breast cancer cell lines, and found that on average, a 2-fold change in DNA copy number was associated with a corresponding 1.5-fold change in mRNA levels. In summary, the evidence supports the Appellants' position that gene amplification is more likely than not predictive of increased mRNA and polypeptide levels.

Second, the Declaration of Dr. Paul Polakis (made of record in Appellants' Response filed June 28, 2004), principal investigator of the Tumor Antigen Project of Genentech, Inc., the assignee of the present application, explains that in the course of Dr. Polakis' research using microarray analysis, he and his co-workers identified approximately 200 gene transcripts that are present in human tumor cells at significantly higher levels than in corresponding normal human cells. Appellants submit that Dr. Polakis' Declaration was presented to support the position that there is a correlation between mRNA levels and polypeptide levels, the correlation between gene amplification and mRNA levels having already been established by the data shown in the Ornoft *et al.*, Hyman *et al.*, and Pollack *et al.* articles. Appellants further emphasize that the opinions expressed in the Polakis Declaration, including in the above quoted statement, are all based on factual findings. For instance, antibodies binding to about 30 of these tumor antigens were prepared, and mRNA and protein levels were compared. In approximately 80% of the cases, the researchers found that increases in the level of a particular mRNA correlated with changes in the level of protein expressed from that mRNA when human tumor cells are compared with their corresponding normal cells. Therefore, Dr. Polakis' research, which is referenced in his Declaration, shows that, in general, there is a correlation between increased mRNA and polypeptide levels.

Appellants further note that the sale of gene expression chips to measure mRNA levels is a highly successful business, with a company such as Affymetrix recording 168.3 million dollars in sales of their GeneChip® arrays in 2004. Clearly, the research community believe that the

information obtained from these chips is useful (i.e., that it is more likely than not that the results are informative of protein levels).

Taken together, all of the submitted evidence supports the Appellants' position that, in the majority of amplified genes, increased gene amplification levels, more likely than not, predict increased mRNA and polypeptide levels, which clearly meets the utility standards described above. Hence, one of skill in the art would reasonably expect that, based on the gene amplification data of the PRO317 gene, the PRO317 polypeptide is concomitantly overexpressed in the lung or colon tumors studied as well.

Appellants further submit that, even if there were no correlation between gene amplification and increased mRNA/protein expression, (which Appellants expressly do not concede), a polypeptide encoded by an amplified gene in cancer would **still** have a specific, substantial, and credible utility as explained below. As the Declaration of Dr. Avi Ashkenazi (submitted with Appellants' Response filed June 28, 2004) explains:

> "even when amplification of a cancer marker gene does not result in significant over-expression of the corresponding gene product, this very absence of gene product over-expression still provides significant information for cancer diagnosis and treatment."

Thus, even if over-expression of the gene product does not parallel gene amplification in certain tumor types, parallel monitoring of gene amplification and gene product over-expression enables more accurate tumor classification and hence better determination of suitable therapy. In addition, absence of over-expression is crucial information for the practicing clinician. If a gene is amplified in a tumor, but the corresponding gene product is not over-expressed, the clinician will decide not to treat a patient with agents that target that gene product. This not only saves money, but also has the benefit that the patient can avoid exposure to the side effects associated with such agents.

This utility is further supported by the teachings of the article by Hanna and Mornin. (Pathology Associates Medical Laboratories, August (1999), submitted with the Response filed June 28, 2004). The article teaches that the HER-2/neu gene has been shown to be amplified and/or over-expressed in 10%-30% of invasive breast cancers and in 40%-60% of intraductal breast carcinomas. Further, the article teaches that diagnosis of breast cancer includes testing both the amplification of the HER-2/neu gene (by FISH) as well as the over-expression of the HER-2/neu gene product (by IHC). Even when the protein is not over-expressed, the assay

relying on both tests leads to a more accurate classification of the cancer and a more effective treatment of it.

Thus, based on the asserted utility for PRO317 in the diagnosis of selected lung or colon tumors, the reduction to practice of the instantly claimed protein sequence of SEQ ID NO: 114 in the present application (also see page 15, page 41 and pages 103-104, page 133, line 16 to page 135, line 18 and also pages 16-17 and page 133, line 16 to page 135, line 18 and Example 18), the step-by-step preparation of chimeric PRO polypeptides, including those wherein the heterologous polypeptide is an epitope tag or an Fc region of an immunoglobulin (page 74, lines 23 to page 75, line 5), the description of the expression of PRO polypeptides in various host cells, including *E. coli*, mammalian cells, yeast and Baculovirus-infected insect cells at least in Examples 53-56, pages 192-199, the disclosure of the step-by-step protocol for the preparation, isolation and detection of monoclonal, polyclonal and other types of antibodies against the PRO317 protein in the specification (monoclonal and polyclonal antibodies at page 139, line 32, to page 141, line 13; humanized antibodies at page 141, line 15, to page 142, line 16; antibody fragments at page 143, line 8 onwards; labeled antibodies at pages 144-145, line 16 onwards and page 146, line 33 to page 147, line 6) and the disclosure of the gene amplification assay in Example 92, the skilled artisan would know exactly how to make and use the claimed antibodies for the diagnosis of lung and colon cancers. Appellants submit that based on the detailed information presented in the specification and the advanced state of the art in oncology, the skilled artisan would have found such testing routine and not 'undue.'

Therefore, since the instantly claimed invention is supported by either a credible, specific and substantial asserted utility or a well-established utility based on the PCT/US98/18824 specification and since it also clearly teaches one skilled in the art "how to make and use" the claimed invention without undue experimentation, Appellants respectfully request reconsideration and reversal of the determination of priority for Claims 44-46 and 49-51.

ISSUE 2. Claims 44-46 and 49-51 are not anticipated by Ruben *et al.*, WO99/09198

Claims 44-46 and 49-51 remain rejected under 35 U.S.C. §102(a) as being anticipated by Ruben *et al.*, WO99/09198 (February 1999).

For the reasons discussed above, Appellants maintain that they are entitled to an effective filing date of September 10, 1998 based on a properly claimed priority to International

application PCT/US98/18824. Therefore, Ruben *et al.*, WO99/09198, dated February 1999, is not prior art and the instant claims are not anticipated by Ruben *et al.* Accordingly, this rejection under 35 U.S.C. §102(a) based on Ruben *et al.* should be withdrawn.

CONCLUSION

For the reasons given above, Appellants submit that present specification and the specification of PCT/US98/18824 filed on September 10, 1998 clearly describes and provides at least one patentable utility for the instantly claimed invention. Moreover, it is respectfully submitted that the present specification clearly teaches "how to use" the presently claimed polypeptide based upon this disclosed patentable utility. Accordingly, Ruben *et al.*, WO99/09198 is not prior art. As such, Appellants respectfully request reconsideration and reversal of the outstanding rejection of claims 44-46 and 49-51.

The Commissioner is authorized to charge any fees which may be required, including extension fees, or credit any overpayment to Deposit Account No. **08-1641** (referencing Attorney's Docket No. **39780-1618 P2C17**.

Respectfully submitted,

(Panpan Gao 43,626) on behalf of Daphne Reddy

Date: March 15, 2006

Daphne Reddy
Reg. No. 53,507

HELLER EHRMAN LLP
275 Middlefield Road
Menlo Park, California 94025-3506
Telephone: (650) 324-7000
Facsimile: (650) 324-0638

IX. CLAIMS APPENDIX

Claims on Appeal

44. An isolated polypeptide comprising:

(a) the amino acid sequence of the polypeptide of SEQ ID NO: 114;

(b) the amino acid sequence of the polypeptide of SEQ ID NO: 114, lacking its associated signal peptide;

(c) the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 209367,

wherein, the nucleic acid encoding said polypeptide is amplified in lung or colon tumors.

45. The isolated polypeptide of Claim 44 comprising the amino acid sequence of the polypeptide of SEQ ID NO: 114.

46. The isolated polypeptide of Claim 44 comprising the amino acid sequence of the polypeptide of SEQ ID NO: 114, lacking its associated signal peptide.

49. The isolated polypeptide of Claim 44 comprising the amino acid sequence of the polypeptide encoded by the full-length coding sequence of the cDNA deposited under ATCC accession number 209367.

50. A chimeric polypeptide comprising a polypeptide according to Claim 44 fused to a heterologous polypeptide.

51. The chimeric polypeptide of Claim 50, wherein said heterologous polypeptide is an epitope tag or an Fc region of an immunoglobulin.

X. EVIDENCE APPENDIX

1. Ruben *et al.*, WO 99/09198 (February 1999).

2. Pennica *et al.*, "WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors," *Proc. Natl. Acad. Sci. USA* **95**:14717-14722 (1998).

3. Declaration of Paul Polakis, Ph.D. under 35 C.F.R. 1.132.

4. Declaration of Avi Ashkenazi, Ph.D. under 35 C.F.R. 1.132, with attached Exhibit A (Curriculum Vitae).

5. Orntoft, T.F., et al., "Genome-wide Study of Gene Copy Numbers, Transcripts, and Protein Levels in Pairs of Non-Invasive and Invasive Human Transitional Cell Carcinomas," *Molecular & Cellular Proteomics* **1**:37-45 (2002).

6. Hyman, E., et al., "Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer," *Cancer Research* **62**:6240-6245 (2002).

7. Pollack, J.R., et al., "Microarray Analysis Reveals a Major Direct Role of DNA Copy Number Alteration in the Transcriptional Program of Human Breast Tumors," *Proc. Natl. Acad. Sci. USA* **99**:12963-12968 (2002).

8. Hanna et al., "HER-2/neu Breast Cancer Predictive Testing," Pathology Associates Medical Laboratories (1999).

9. Haynes *et al.*, "Proteome analysis: Biological assay or data archive?" *Electrophoresis* **19**:1862-1871 (1996).

10. Hancock *et al.*, "Do we have enough biomarkers" *J. Proteome Res.* **3**(4): 685 (2004).

Item 1 was made of record by the Examiner in the Office Action mailed September 9, 2003.

Item 2 was made of record by the Examiner in the Office Action mailed December 29, 2003.

Items 3-8 were submitted with Appellants' Response filed June 28, 2004, and were considered by the Examiner as indicated in the Office action mailed September 20, 2004.

Items 9-10 were made of record by the Examiner in the Office Action mailed September 20, 2004.

XI. RELATED PROCEEDINGS APPENDIX

None- no decision rendered by a Court or the Board in any related proceedings identified above.

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 14717–14722, December 1998
Cell Biology, Medical Sciences

WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors

DIANE PENNICA*†, TODD A. SWANSON*, JAMES W. WELSH*, MARGARET A. ROY‡, DAVID A. LAWRENCE*, JAMES LEE‡, JENNIFER BRUSH‡, LISA A. TANEYHILL§, BETHANNE DEUEL‡, MICHAEL LEW¶, COLIN WATANABE‖, ROBERT L. COHEN*, MONA F. MELHEM**, GENE G. FINLEY**, PHIL QUIRKE††, AUDREY D. GODDARD‡, KENNETH J. HILLAN¶, AUSTIN L. GURNEY‡, DAVID BOTSTEIN‡,‡‡, AND ARNOLD J. LEVINE§

Departments of *Molecular Oncology, ‡Molecular Biology, ‖Scientific Computing, and ¶Pathology, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080; **University of Pittsburgh School of Medicine, Veterans Administration Medical Center, Pittsburgh, PA 15240; ††University of Leeds, Leeds, LS29JT United Kingdom; ‡‡Department of Genetics, Stanford University, Palo Alto, CA 94305; and §Department of Molecular Biology, Princeton University, Princeton, NJ 08544

Contributed by David Botstein and Arnold J. Levine, October 21, 1998

ABSTRACT **Wnt family members are critical to many developmental processes, and components of the Wnt signaling pathway have been linked to tumorigenesis in familial and sporadic colon carcinomas. Here we report the identification of two genes, *WISP-1* and *WISP-2*, that are up-regulated in the mouse mammary epithelial cell line C57MG transformed by Wnt-1, but not by Wnt-4. Together with a third related gene, *WISP-3*, these proteins define a subfamily of the connective tissue growth factor family. Two distinct systems demonstrated *WISP* induction to be associated with the expression of Wnt-1. These included (*i*) C57MG cells infected with a Wnt-1 retroviral vector or expressing Wnt-1 under the control of a tetracyline repressible promoter, and (*ii*) Wnt-1 transgenic mice. The *WISP-1* gene was localized to human chromosome 8q24.1–8q24.3. *WISP-1* genomic DNA was amplified in colon cancer cell lines and in human colon tumors and its RNA overexpressed (2- to >30-fold) in 84% of the tumors examined compared with patient-matched normal mucosa. *WISP-3* mapped to chromosome 6q22–6q23 and also was overexpressed (4- to >40-fold) in 63% of the colon tumors analyzed. In contrast, *WISP-2* mapped to human chromosome 20q12–20q13 and its DNA was amplified, but RNA expression was reduced (2- to >30-fold) in 79% of the tumors. These results suggest that the *WISP* genes may be downstream of Wnt-1 signaling and that aberrant levels of *WISP* expression in colon cancer may play a role in colon tumorigenesis.**

Wnt-1 is a member of an expanding family of cysteine-rich, glycosylated signaling proteins that mediate diverse developmental processes such as the control of cell proliferation, adhesion, cell polarity, and the establishment of cell fates (1, 2). Wnt-1 originally was identified as an oncogene activated by the insertion of mouse mammary tumor virus in virus-induced mammary adenocarcinomas (3, 4). Although Wnt-1 is not expressed in the normal mammary gland, expression of Wnt-1 in transgenic mice causes mammary tumors (5).

In mammalian cells, Wnt family members initiate signaling by binding to the seven-transmembrane spanning Frizzled receptors and recruiting the cytoplasmic protein Dishevelled (Dsh) to the cell membrane (1, 2, 6). Dsh then inhibits the kinase activity of the normally constitutively active glycogen synthase kinase-3β (GSK-3β) resulting in an increase in β-catenin levels. Stabilized β-catenin interacts with the transcription factor TCF/Lef1, forming a complex that appears in the nucleus and binds TCF/Lef1 target DNA elements to activate transcription (7, 8). Other experiments suggest that the adenomatous polyposis coli (APC) tumor suppressor gene also plays an important role in Wnt signaling by regulating β-catenin levels (9). APC is phosphorylated by GSK-3β, binds to β-catenin, and facilitates its degradation. Mutations in either APC or β-catenin have been associated with colon carcinomas and melanomas, suggesting these mutations contribute to the development of these types of cancer, implicating the Wnt pathway in tumorigenesis (1).

Although much has been learned about the Wnt signaling pathway over the past several years, only a few of the transcriptionally activated downstream components activated by Wnt have been characterized. Those that have been described cannot account for all of the diverse functions attributed to Wnt signaling. Among the candidate Wnt target genes are those encoding the nodal-related 3 gene, *Xnr3*, a member of the transforming growth factor (TGF)-β superfamily, and the homeobox genes, *engrailed*, *goosecoid*, *twin* (*Xtwn*), and *siamois* (2). A recent report also identifies c-*myc* as a target gene of the Wnt signaling pathway (10).

To identify additional downstream genes in the Wnt signaling pathway that are relevant to the transformed cell phenotype, we used a PCR-based cDNA subtraction strategy, suppression subtractive hybridization (SSH) (11), using RNA isolated from C57MG mouse mammary epithelial cells and C57MG cells stably transformed by a Wnt-1 retrovirus. Overexpression of Wnt-1 in this cell line is sufficient to induce a partially transformed phenotype, characterized by elongated and refractile cells that lose contact inhibition and form a multilayered array (12, 13). We reasoned that genes differentially expressed between these two cell lines might contribute to the transformed phenotype.

In this paper, we describe the cloning and characterization of two genes up-regulated in Wnt-1 transformed cells, *WISP-1* and *WISP-2*, and a third related gene, *WISP-3*. The *WISP* genes are members of the CCN family of growth factors, which includes connective tissue growth factor (CTGF), Cyr61, and *nov*, a family not previously linked to Wnt signaling.

MATERIALS AND METHODS

SSH. SSH was performed by using the PCR-Select cDNA Subtraction Kit (CLONTECH). Tester double-stranded

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: TGF, transforming growth factor; CTGF, connective tissue growth factor; SSH, suppression subtractive hybridization; VWC, von Willebrand factor type C module.

Data deposition: The sequences reported in this paper have been deposited in the Genbank database (accession nos. AF100777, AF100778, AF100779, AF100780, and AF100781).

†To whom reprint requests should be addressed. e-mail: diane@gene.com.

cDNA was synthesized from 2 μg of poly(A)$^+$ RNA isolated from the C57MG/Wnt-1 cell line and driver cDNA from 2 μg of poly(A)$^+$ RNA from the parent C57MG cells. The subtracted cDNA library was subcloned into a pGEM-T vector for further analysis.

cDNA Library Screening. Clones encoding full-length mouse *WISP-1* were isolated by screening a λgt10 mouse embryo cDNA library (CLONTECH) with a 70-bp probe from the original partial clone 568 sequence corresponding to amino acids 128–169. Clones encoding full-length human *WISP-1* were isolated by screening λgt10 lung and fetal kidney cDNA libraries with the same probe at low stringency. Clones encoding full-length mouse and human *WISP-2* were isolated by screening a C57MG/Wnt-1 or human fetal lung cDNA library with a probe corresponding to nucleotides 1463–1512. Full-length cDNAs encoding *WISP-3* were cloned from human bone marrow and fetal kidney libraries.

Expression of Human *WISP* RNA. PCR amplification of first-strand cDNA was performed with human Multiple Tissue cDNA panels (CLONTECH) and 300 μM of each dNTP at 94°C for 1 sec, 62°C for 30 sec, 72°C for 1 min, for 22–32 cycles. *WISP* and glyceraldehyde-3-phosphate dehydrogenase primer sequences are available on request.

***In Situ* Hybridization.** ^{33}P-labeled sense and antisense riboprobes were transcribed from an 897-bp PCR product corresponding to nucleotides 601–1440 of mouse *WISP-1* or a 294-bp PCR product corresponding to nucleotides 82–375 of mouse *WISP-2*. All tissues were processed as described (40).

Radiation Hybrid Mapping. Genomic DNA from each hybrid in the Stanford G3 and Genebridge4 Radiation Hybrid Panels (Research Genetics, Huntsville, AL) and human and hamster control DNAs were PCR-amplified, and the results were submitted to the Stanford or Massachusetts Institute of Technology web servers.

Cell Lines, Tumors, and Mucosa Specimens. Tissue specimens were obtained from the Department of Pathology (University of Pittsburgh) for patients undergoing colon resection and from the University of Leeds, United Kingdom. Genomic DNA was isolated (Qiagen) from the pooled blood of 10 normal human donors, surgical specimens, and the following ATCC human cell lines: SW480, COLO 320DM, HT-29, WiDr, and SW403 (colon adenocarcinomas), SW620 (lymph node metastasis, colon adenocarcinoma), HCT 116 (colon carcinoma), SK-CO-1 (colon adenocarcinoma, ascites), and HM7 (a variant of ATCC colon adenocarcinoma cell line LS 174T). DNA concentration was determined by using Hoechst dye 33258 intercalation fluorimetry. Total RNA was prepared by homogenization in 7 M GuSCN followed by centrifugation over CsCl cushions or prepared by using RNAzol.

Gene Amplification and RNA Expression Analysis. Relative gene amplification and RNA expression of *WISP*s and c-*myc* in the cell lines, colorectal tumors, and normal mucosa were determined by quantitative PCR. Gene-specific primers and fluorogenic probes (sequences available on request) were designed and used to amplify and quantitate the genes. The relative gene copy number was derived by using the formula $2^{(\Delta ct)}$ where ΔCt represents the difference in amplification cycles required to detect the *WISP* genes in peripheral blood lymphocyte DNA compared with colon tumor DNA or colon tumor RNA compared with normal mucosal RNA. The ∂-method was used for calculation of the SE of the gene copy number or RNA expression level. The *WISP*-specific signal was normalized to that of the glyceraldehyde-3-phosphate dehydrogenase housekeeping gene. All TaqMan assay reagents were obtained from Perkin–Elmer Applied Biosystems.

RESULTS

Isolation of *WISP-1* and *WISP-2* by SSH. To identify Wnt-1-inducible genes, we used the technique of SSH using the mouse mammary epithelial cell line C57MG and C57MG cells that stably express Wnt-1 (11). Candidate differentially expressed cDNAs (1,384 total) were sequenced. Thirty-nine percent of the sequences matched known genes or homologues, 32% matched expressed sequence tags, and 29% had no match. To confirm that the transcript was differentially expressed, semiquantitative reverse transcription–PCR and Northern analysis were performed by using mRNA from the C57MG and C57MG/Wnt-1 cells.

Two of the cDNAs, *WISP-1* and *WISP-2*, were differentially expressed, being induced in the C57MG/Wnt-1 cell line, but not in the parent C57MG cells or C57MG cells overexpressing Wnt-4 (Fig. 1 *A* and *B*). Wnt-4, unlike Wnt-1, does not induce the morphological transformation of C57MG cells and has no effect on β-catenin levels (13, 14). Expression of *WISP-1* was up-regulated approximately 3-fold in the C57MG/Wnt-1 cell line and *WISP-2* by approximately 5-fold by both Northern analysis and reverse transcription–PCR.

An independent, but similar, system was used to examine *WISP* expression after Wnt-1 induction. C57MG cells expressing the *Wnt-1* gene under the control of a tetracycline-repressible promoter produce low amounts of Wnt-1 in the repressed state but show a strong induction of *Wnt-1* mRNA and protein within 24 hr after tetracycline removal (8). The levels of Wnt-1 and *WISP* RNA isolated from these cells at various times after tetracycline removal were assessed by quantitative PCR. Strong induction of Wnt-1 mRNA was seen as early as 10 hr after tetracycline removal. Induction of *WISP* mRNA (2- to 6-fold) was seen at 48 and 72 hr (data not shown). These data support our previous observations that show that *WISP* induction is correlated with Wnt-1 expression. Because the induction is slow, occurring after approximately 48 hr, the induction of *WISP*s may be an indirect response to Wnt-1 signaling.

cDNA clones of human *WISP-1* were isolated and the sequence compared with mouse *WISP-1*. The cDNA sequences of mouse and human *WISP-1* were 1,766 and 2,830 bp in length, respectively, and encode proteins of 367 aa, with predicted relative molecular masses of ≈40,000 (M_r 40 K). Both have hydrophobic N-terminal signal sequences, 38 conserved cysteine residues, and four potential N-linked glycosylation sites and are 84% identical (Fig. 2*A*).

Full-length cDNA clones of mouse and human *WISP-2* were 1,734 and 1,293 bp in length, respectively, and encode proteins of 251 and 250 aa, respectively, with predicted relative molecular masses of ≈27,000 (M_r 27 K) (Fig. 2*B*). Mouse and human *WISP-2* are 73% identical. Human *WISP-2* has no potential N-linked glycosylation sites, and mouse *WISP-2* has one at

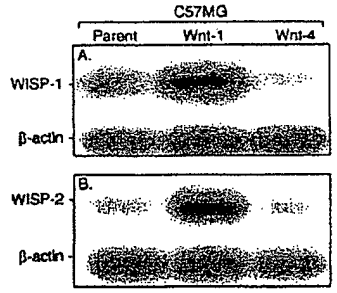


FIG. 1. *WISP-1* and *WISP-2* are induced by Wnt-1, but not Wnt-4, expression in C57MG cells. Northern analysis of *WISP-1* (*A*) and *WISP-2* (*B*) expression in C57MG, C57MG/Wnt-1, and C57MG/Wnt-4 cells. Poly(A)$^+$ RNA (2 μg) was subjected to Northern blot analysis and hybridized with a 70-bp mouse *WISP-1*-specific probe (amino acids 278–300) or a 190-bp *WISP-2*-specific probe (nucleotides 1438–1627) in the 3′ untranslated region. Blots were rehybridized with human β-actin probe.

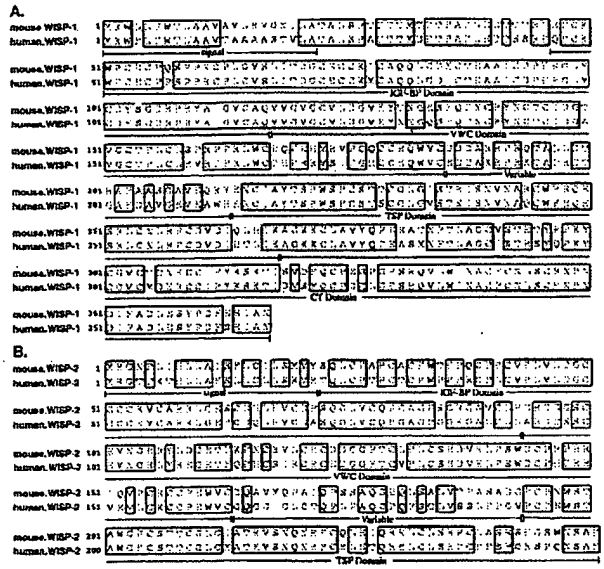


FIG. 2. Encoded amino acid sequence alignment of mouse and human *WISP-1* (*A*) and mouse and human *WISP-2* (*B*). The potential signal sequence, insulin-like growth factor-binding protein (IGF-BP), VWC, thrombospondin (TSP), and C-terminal (CT) domains are underlined.

position 197. *WISP-2* has 28 cysteine residues that are conserved among the 38 cysteines found in *WISP-1*.

Identification of *WISP-3*. To search for related proteins, we screened expressed sequence tag (EST) databases with the WISP-1 protein sequence and identified several ESTs as potentially related sequences. We identified a homologous protein that we have called WISP-3. A full-length human *WISP-3* cDNA of 1,371 bp was isolated corresponding to those ESTs that encode a 354-aa protein with a predicted molecular mass of 39,293. WISP-3 has two potential N-linked glycosylation sites and 36 cysteine residues. An alignment of the three human WISP proteins shows that WISP-1 and WISP-3 are the most similar (42% identity), whereas WISP-2 has 37% identity with WISP-1 and 32% identity with WISP-3 (Fig. 3*A*).

***WISPs* Are Homologous to the CTGF Family of Proteins.** Human *WISP-1*, *WISP-2*, and *WISP-3* are novel sequences; however, mouse *WISP-1* is the same as the recently identified *Elm1* gene. *Elm1* is expressed in low, but not high, metastatic mouse melanoma cells, and suppresses the *in vivo* growth and metastatic potential of K-1735 mouse melanoma cells (15). Human and mouse *WISP-2* are homologous to the recently described rat gene, *rCop-1* (16). Significant homology (36–44%) was seen to the CCN family of growth factors. This family includes three members, CTGF, Cyr61, and the protooncogene *nov*. CTGF is a chemotactic and mitogenic factor for fibroblasts that is implicated in wound healing and fibrotic disorders and is induced by TGF-β (17). Cyr61 is an extracellular matrix signaling molecule that promotes cell adhesion, proliferation, migration, angiogenesis, and tumor growth (18, 19). *nov* (nephroblastoma overexpressed) is an immediate early gene associated with quiescence and found altered in Wilms tumors (20). The proteins of the CCN family share functional, but not sequence, similarity to Wnt-1. All are secreted, cysteine-rich heparin binding glycoproteins that associate with the cell surface and extracellular matrix.

WISP proteins exhibit the modular architecture of the CCN family, characterized by four conserved cysteine-rich domains (Fig. 3*B*) (21). The N-terminal domain, which includes the first 12 cysteine residues, contains a consensus sequence (GCGCCXXC) conserved in most insulin-like growth factor (IGF)-binding proteins (BP). This sequence is conserved in WISP-2 and WISP-3, whereas WISP-1 has a glutamine in the third position instead of a glycine. CTGF recently has been shown to specifically bind IGF (22) and a truncated *nov* protein lacking the IGF-BP domain is oncogenic (23). The von Willebrand factor type C module (VWC), also found in certain collagens and mucins, covers the next 10 cysteine residues, and is thought to participate in protein complex formation and oligomerization (24). The VWC domain of WISP-3 differs from all CCN family members described previously, in that it contains only six of the 10 cysteine residues (Fig. 3 *A* and *B*). A short variable region follows the VWC domain. The third module, the thrombospondin (TSP) domain is involved in binding to sulfated glycoconjugates and contains six cysteine residues and a conserved WSxCSxxCG motif first identified in thrombospondin (25). The C-terminal (CT) module containing the remaining 10 cysteines is thought to be involved in dimerization and receptor binding (26). The CT domain is present in all CCN family members described to date but is absent in WISP-2 (Fig. 3 *A* and *B*). The existence of a putative signal sequence and the absence of a transmembrane domain suggest that WISPs are secreted proteins, an observation supported by an analysis of their expression and secretion from mammalian cell and baculovirus cultures (data not shown).

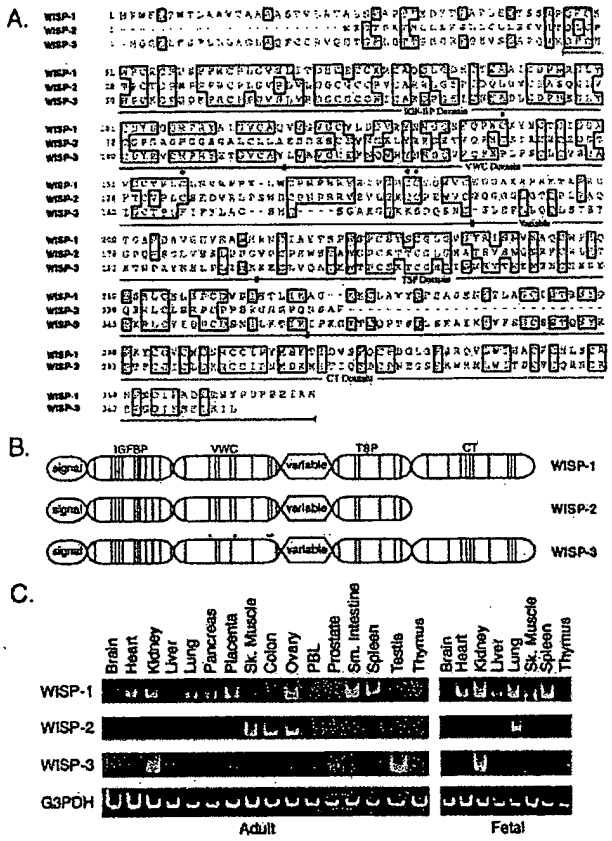


FIG. 3. (*A*) Encoded amino acid sequence alignment of human WISPs. The cysteine residues of WISP-1 and WISP-2 that are not present in WISP-3 are indicated with a dot. (*B*) Schematic representation of the WISP proteins showing the domain structure and cysteine residues (vertical lines). The four cysteine residues in the VWC domain that are absent in WISP-3 are indicated with a dot. (*C*) Expression of *WISP* mRNA in human tissues. PCR was performed on human multiple-tissue cDNA panels (CLONTECH) from the indicated adult and fetal tissues.

Expression of *WISP* mRNA in Human Tissues. Tissue-specific expression of human *WISP*s was characterized by PCR

analysis on adult and fetal multiple tissue cDNA panels. *WISP-1* expression was seen in the adult heart, kidney, lung, pancreas, placenta, ovary, small intestine, and spleen (Fig. 3*C*). Little or no expression was detected in the brain, liver, skeletal muscle, colon, peripheral blood leukocytes, prostate, testis, or thymus. *WISP-2* had a more restricted tissue expression and was detected in adult skeletal muscle, colon, ovary, and fetal lung. Predominant expression of *WISP-3* was seen in adult kidney and testis and fetal kidney. Lower levels of *WISP-3* expression were detected in placenta, ovary, prostate, and small intestine.

***In Situ* Localization of *WISP-1* and *WISP-2*.** Expression of *WISP-1* and *WISP-2* was assessed by *in situ* hybridization in mammary tumors from Wnt-1 transgenic mice. Strong expression of *WISP-1* was observed in stromal fibroblasts lying within the fibrovascular tumor stroma (Fig. 4 *A–D*). However, low-level *WISP-1* expression also was observed focally within tumor cells (data not shown). No expression was observed in normal breast. Like *WISP-1*, *WISP-2* expression also was seen in the tumor stroma in breast tumors from Wnt-1 transgenic animals (Fig. 4 *E–H*). However, *WISP-2* expression in the stroma was in spindle-shaped cells adjacent to capillary vessels, whereas

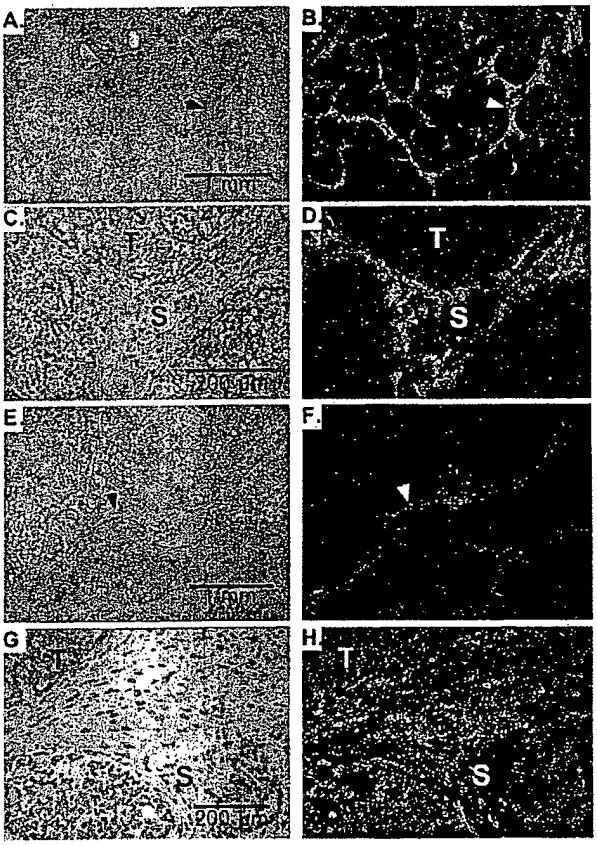


FIG. 4. (*A*, *C*, *E*, and *G*) Representative hematoxylin/eosin-stained images from breast tumors in Wnt-1 transgenic mice. The corresponding dark-field images showing *WISP-1* expression are shown in *B* and *D*. The tumor is a moderately well-differentiated adenocarcinoma showing evidence of adenoid cystic change. At low power (*A* and *B*), expression of *WISP-1* is seen in the delicate branching fibrovascular tumor stroma (arrowhead). At higher magnification, expression is seen in the stromal(s) fibroblasts (*C* and *D*), and tumor cells are negative. Focal expression of *WISP-1*, however, was observed in tumor cells in some areas. Images of *WISP-2* expression are shown in *E–H*. At low power (*E* and *F*), expression of WISP-2 is seen in cells lying within the fibrovascular tumor stroma. At higher magnification, these cells appeared to be adjacent to capillary vessels whereas tumor cells are negative (*G* and *H*).

the predominant cell type expressing *WISP-1* was the stromal fibroblasts.

Chromosome Localization of the *WISP* Genes. The chromosomal location of the human *WISP* genes was determined by radiation hybrid mapping panels. *WISP-1* is approximately 3.48 cR from the meiotic marker AFM259xc5 [logarithm of odds (lod) score 16.31] on chromosome 8q24.1 to 8q24.3, in the same region as the human locus of the *nov*H family member (27) and roughly 4 Mbs distal to c-*myc* (28). Preliminary fine mapping indicates that *WISP-1* is located near D8S1712 STS. *WISP-2* is linked to the marker SHGC-33922 (lod = 1,000) on chromosome 20q12–20q13.1. Human *WISP-3* mapped to chromosome 6q22–6q23 and is linked to the marker AFM211ze5 (lod = 1,000). *WISP-3* is approximately 18 Mbs proximal to CTGF and 23 Mbs proximal to the human cellular oncogene *MYB* (27, 29).

Amplification and Aberrant Expression of *WISPs* in Human Colon Tumors. Amplification of protooncogenes is seen in many human tumors and has etiological and prognostic significance. For example, in a variety of tumor types, c-*myc* amplification has been associated with malignant progression and poor prognosis (30). Because *WISP-1* resides in the same general chromosomal location (8q24) as c-*myc*, we asked whether it was a target of gene amplification, and, if so, whether this amplification was independent of the c-*myc* locus. Genomic DNA from human colon cancer cell lines was assessed by quantitative PCR and Southern blot analysis. (Fig. 5 *A* and *B*). Both methods detected similar degrees of *WISP-1* amplification. Most cell lines showed significant (2- to 4-fold) amplification, with the HT-29 and WiDr cell lines demonstrating an 8-fold increase. Significantly, the pattern of amplification observed did not correlate with that observed for c-*myc*, indicating that the c-*myc* gene is not part of the amplicon that involves the *WISP-1* locus.

We next examined whether the *WISP* genes were amplified in a panel of 25 primary human colon adenocarcinomas. The relative *WISP* gene copy number in each colon tumor DNA was compared with pooled normal DNA from 10 donors by quantitative PCR (Fig. 6). The copy number of *WISP-1* and *WISP-2* was significantly greater than one, approximately 2-fold for *WISP-1* in about 60% of the tumors and 2- to 4-fold for *WISP-2* in 92% of the tumors ($P < 0.001$ for each). The copy number for *WISP-3* was indistinguishable from one ($P = 0.166$). In addition, the copy number of *WISP-2* was significantly higher than that of *WISP-1* ($P < 0.001$).

The levels of *WISP* transcripts in RNA isolated from 19 adenocarcinomas and their matched normal mucosa were

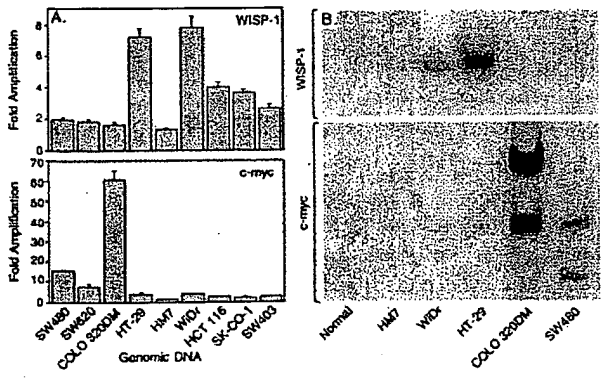


FIG. 5. Amplification of *WISP-1* genomic DNA in colon cancer cell lines. (*A*) Amplification in cell line DNA was determined by quantitative PCR. (*B*) Southern blots containing genomic DNA (10 μg) digested with *Eco*RI (*WISP-1*) or *Xba*I (c-*myc*) were hybridized with a 100-bp human *WISP-1* probe (amino acids 186–219) or a human c-*myc* probe (located at bp 1901–2000). The *WISP* and *myc* genes are detected in normal human genomic DNA after a longer film exposure.

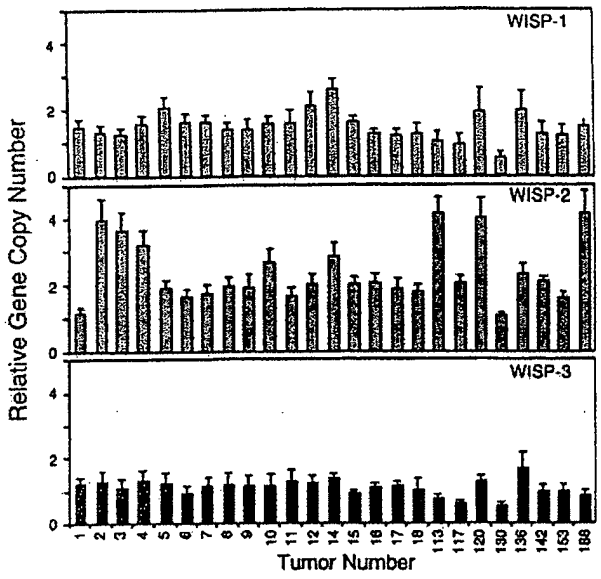


FIG. 6. Genomic amplification of *WISP* genes in human colon tumors. The relative gene copy number of the *WISP* genes in 25 adenocarcinomas was assayed by quantitative PCR, by comparing DNA from primary human tumors with pooled DNA from 10 healthy donors. The data are means ± SEM from one experiment done in triplicate. The experiment was repeated at least three times.

assessed by quantitative PCR (Fig. 7). The level of *WISP-1* RNA present in tumor tissue varied but was significantly increased (2- to >25-fold) in 84% (16/19) of the human colon tumors examined compared with normal adjacent mucosa. Four of 19 tumors showed greater than 10-fold overexpression. In contrast, in 79% (15/19) of the tumors examined, *WISP-2* RNA expression was significantly lower in the tumor than the mucosa. Similar to *WISP-1*, *WISP-3* RNA was overexpressed in 63% (12/19) of the colon tumors compared with the normal

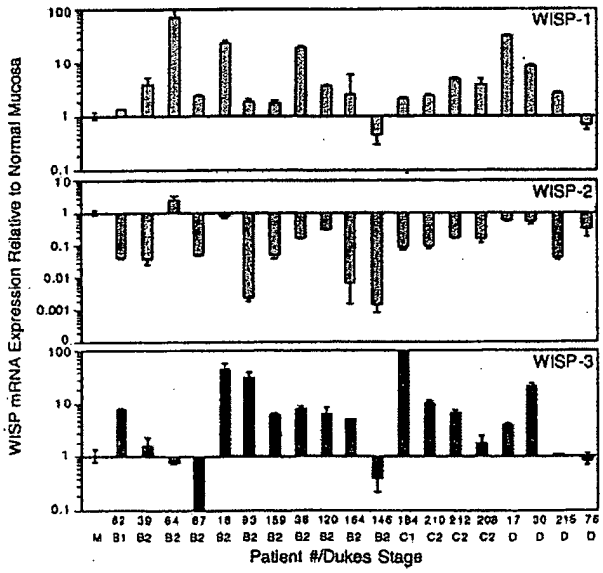


FIG. 7. *WISP* RNA expression in primary human colon tumors relative to expression in normal mucosa from the same patient. Expression of *WISP* mRNA in 19 adenocarcinomas was assayed by quantitative PCR. The Dukes stage of the tumor is listed under the sample number. The data are means ± SEM from one experiment done in triplicate. The experiment was repeated at least twice.

mucosa. The amount of overexpression of *WISP-3* ranged from 4- to >40-fold.

DISCUSSION

One approach to understanding the molecular basis of cancer is to identify differences in gene expression between cancer cells and normal cells. Strategies based on assumptions that steady-state mRNA levels will differ between normal and malignant cells have been used to clone differentially expressed genes (31). We have used a PCR-based selection strategy, SSH, to identify genes selectively expressed in C57MG mouse mammary epithelial cells transformed by Wnt-1.

Three of the genes isolated, *WISP-1*, *WISP-2*, and *WISP-3*, are members of the CCN family of growth factors, which includes CTGF, Cyr61, and *nov*, a family not previously linked to Wnt signaling.

Two independent experimental systems demonstrated that *WISP* induction was associated with the expression of Wnt-1. The first was C57MG cells infected with a Wnt-1 retroviral vector or C57MG cells expressing Wnt-1 under the control of a tetracyline-repressible promoter, and the second was in Wnt-1 transgenic mice, where breast tissue expresses Wnt-1, whereas normal breast tissue does not. No *WISP* RNA expression was detected in mammary tumors induced by polyoma virus middle T antigen (data not shown). These data suggest a link between Wnt-1 and *WISP*s in that in these two situations, *WISP* induction was correlated with Wnt-1 expression.

It is not clear whether the *WISP*s are directly or indirectly induced by the downstream components of the Wnt-1 signaling pathway (i.e., β-catenin-TCF-1/Lef1). The increased levels of *WISP* RNA were measured in Wnt-1-transformed cells, hours or days after Wnt-1 transformation. Thus, *WISP* expression could result from Wnt-1 signaling directly through β-catenin transcription factor regulation or alternatively through Wnt-1 signaling turning on a transcription factor, which in turn regulates *WISP*s.

The *WISP*s define an additional subfamily of the CCN family of growth factors. One striking difference observed in the protein sequence of WISP-2 is the absence of a CT domain, which is present in CTGF, Cyr61, *nov*, WISP-1, and WISP-3. This domain is thought to be involved in receptor binding and dimerization. Growth factors, such as TGF-β, platelet-derived growth factor, and nerve growth factor, which contain a cystine knot motif exist as dimers (32). It is tempting to speculate that WISP-1 and WISP-3 may exist as dimers, whereas WISP-2 exists as a monomer. If the CT domain is also important for receptor binding, WISP-2 may bind its receptor through a different region of the molecule than the other CCN family members. No specific receptors have been identified for CTGF or *nov*. A recent report has shown that integrin $\alpha_v\beta_3$ serves as an adhesion receptor for Cyr61 (33).

The strong expression of *WISP-1* and *WISP-2* in cells lying within the fibrovascular tumor stroma in breast tumors from Wnt-1 transgenic animals is consistent with previous observations that transcripts for the related CTGF gene are primarily expressed in the fibrous stroma of mammary tumors (34). Epithelial cells are thought to control the proliferation of connective tissue stroma in mammary tumors by a cascade of growth factor signals similar to that controlling connective tissue formation during wound repair. It has been proposed that mammary tumor cells or inflammatory cells at the tumor interstitial interface secrete TGF-β1, which is the stimulus for stromal proliferation (34). TGF-β1 is secreted by a large percentage of malignant breast tumors and may be one of the growth factors that stimulates the production of CTGF and WISPs in the stroma.

It was of interest that *WISP-1* and *WISP-2* expression was observed in the stromal cells that surrounded the tumor cells

(epithelial cells) in the Wnt-1 transgenic mouse sections of breast tissue. This finding suggests that paracrine signaling could occur in which the stromal cells could supply WISP-1 and WISP-2 to regulate tumor cell growth on the WISP extracellular matrix. Stromal cell-derived factors in the extracellular matrix have been postulated to play a role in tumor cell migration and proliferation (35). The localization of *WISP-1* and *WISP-2* in the stromal cells of breast tumors supports this paracrine model.

An analysis of *WISP-1* gene amplification and expression in human colon tumors showed a correlation between DNA amplification and overexpression, whereas overexpression of *WISP-3* RNA was seen in the absence of DNA amplification. In contrast, *WISP-2* DNA was amplified in the colon tumors, but its mRNA expression was significantly reduced in the majority of tumors compared with the expression in normal colonic mucosa from the same patient. The gene for human *WISP-2* was localized to chromosome 20q12–20q13, at a region frequently amplified and associated with poor prognosis in node negative breast cancer and many colon cancers, suggesting the existence of one or more oncogenes at this locus (36–38). Because the center of the 20q13 amplicon has not yet been identified, it is possible that the apparent amplification observed for *WISP-2* may be caused by another gene in this amplicon.

A recent manuscript on *rCop-1*, the rat orthologue of *WISP-2*, describes the loss of expression of this gene after cell transformation, suggesting it may be a negative regulator of growth in cell lines (16). Although the mechanism by which *WISP-2* RNA expression is down-regulated during malignant transformation is unknown, the reduced expression of *WISP-2* in colon tumors and cell lines suggests that it may function as a tumor suppressor. These results show that the *WISP* genes are aberrantly expressed in colon cancer and suggest that their altered expression may confer selective growth advantage to the tumor.

Members of the Wnt signaling pathway have been implicated in the pathogenesis of colon cancer, breast cancer, and melanoma, including the tumor suppressor gene adenomatous polyposis coli and β-catenin (39). Mutations in specific regions of either gene can cause the stabilization and accumulation of cytoplasmic β-catenin, which presumably contributes to human carcinogenesis through the activation of target genes such as the *WISP*s. Although the mechanism by which Wnt-1 transforms cells and induces tumorigenesis is unknown, the identification of *WISP*s as genes that may be regulated downstream of Wnt-1 in C57MG cells suggests they could be important mediators of Wnt-1 transformation. The amplification and altered expression patterns of the *WISP*s in human colon tumors may indicate an important role for these genes in tumor development.

We thank the DNA synthesis group for oligonucleotide synthesis, T. Baker for technical assistance, P. Dowd for radiation hybrid mapping, K. Willert and R. Nusse for the tet-repressible C57MG/Wnt-1 cells, V. Dixit for discussions, and D. Wood and A. Bruce for artwork.

1. Cadigan, K. M. & Nusse, R. (1997) *Genes Dev.* **11**, 3286–3305.
2. Dale, T. C. (1998) *Biochem. J.* **329**, 209–223.
3. Nusse, R. & Varmus, H. E. (1982) *Cell* **31**, 99–109.
4. van Ooyen, A. & Nusse, R. (1984) *Cell* **39**, 233–240.
5. Tsukamoto, A. S., Grosschedl, R., Guzman, R. C., Parslow, T. & Varmus, H. E. (1988) *Cell* **55**, 619–625.
6. Brown, J. D. & Moon, R. T. (1998) *Curr. Opin. Cell. Biol.* **10**, 182–187.
7. Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O. & Clevers, H. (1996) *Cell* **86**, 391–399.
8. Korinek, V., Barker, N., Willert, K., Molenaar, M., Roose, J., Wagenaar, G., Markman, M., Lamers, W., Destree, O. & Clevers, H. (1998) *Mol. Cell. Biol.* **18**, 1248–1256.
9. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B. & Polakis, P. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 3046–3050.
10. He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B. & Kinzler, K. W. (1998) *Science* **281**, 1509–1512.
11. Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D. & Siebert, P. D. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 6025–6030.
12. Brown, A. M., Wildin, R. S., Prendergast, T. J. & Varmus, H. E. (1986) *Cell* **46**, 1001–1009.
13. Wong, G. T., Gavin, B. J. & McMahon, A. P. (1994) *Mol. Cell. Biol.* **14**, 6278–6286.
14. Shimizu, H., Julius, M. A., Giarre, M., Zheng, Z., Brown, A. M. & Kitajewski, J. (1997) *Cell Growth Differ.* **8**, 1349–1358.
15. Hashimoto, Y., Shindo-Okada, N., Tani, M., Nagamachi, Y., Takeuchi, K., Shiroishi, T., Toma, H. & Yokota, J. (1998) *J. Exp. Med.* **187**, 289–296.
16. Zhang, R., Averboukh, L., Zhu, W., Zhang, H., Jo, H., Dempsey, P. J., Coffey, R. J., Pardee, A. B. & Liang, P. (1998) *Mol. Cell. Biol.* **18**, 6131–6141.
17. Grotendorst, G. R. (1997) *Cytokine Growth Factor Rev.* **8**, 171–179.
18. Kireeva, M. L., Mo, F. E., Yang, G. P. & Lau, L. F. (1996) *Mol. Cell. Biol.* **16**, 1326–1334.
19. Babic, A. M., Kireeva, M. L., Kolesnikova, T. V. & Lau, L. F. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 6355–6360.
20. Martinerie, C., Huff, V., Joubert, I., Badzioch, M., Saunders, G., Strong, L. & Perbal, B. (1994) *Oncogene* **9**, 2729–2732.
21. Bork, P. (1993) *FEBS Lett.* **327**, 125–130.
22. Kim, H. S., Nagalla, S. R., Oh, Y., Wilson, E., Roberts, C. T., Jr. & Rosenfeld, R. G. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 12981–12986.
23. Joliot, V., Martinerie, C., Dambrine, G., Plassiart, G., Brisac, M., Crochet, J. & Perbal, B. (1992) *Mol. Cell. Biol.* **12**, 10–21.
24. Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alevy, Y. G. & Sadler, J. E. (1989) *J. Biol. Chem.* **264**, 19514–19527.
25. Holt, G. D., Pangburn, M. K. & Ginsburg, V. (1990) *J. Biol. Chem.* **265**, 2852–2855.
26. Voorberg, J., Fontijn, R., Calafat, J., Janssen, H., van Mourik, J. A. & Pannekoek, H. (1991) *J. Cell. Biol.* **113**, 195–205.
27. Martinerie, C., Viegas-Pequignot, E., Guenard, I., Dutrillaux, B., Nguyen, V. C., Bernheim, A. & Perbal, B. (1992) *Oncogene* **7**, 2529–2534.
28. Takahashi, E., Hori, T., O'Connell, P., Leppert, M. & White, R. (1991) *Cytogenet. Cell. Genet.* **57**, 109–111.
29. Meese, E., Meltzer, P. S., Witkowski, C. M. & Trent, J. M. (1989) *Genes Chromosomes Cancer* **1**, 88–94.
30. Garte, S. J. (1993) *Crit. Rev. Oncog.* **4**, 435–449.
31. Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B. & Kinzler, K. W. (1997) *Science* **276**, 1268–1272.
32. Sun, P. D. & Davies, D. R. (1995) *Annu. Rev. Biophys. Biomol. Struct.* **24**, 269–291.
33. Kireeva, M. L., Lam, S. C. T. & Lau, L. F. (1998) *J. Biol. Chem.* **273**, 3090–3096.
34. Frazier, K. S. & Grotendorst, G. R. (1997) *Int. J. Biochem. Cell. Biol.* **29**, 153–161.
35. Wernert, N. (1997) *Virchows Arch.* **430**, 433–443.
36. Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Collins, C., Stokke, T., Karhu, R., Kowbel, D., Shadravan, F., Hintz, M., Kuo, W. L., *et al.* (1994) *Cancer Res.* **54**, 4257–4260.
37. Brinkmann, U., Gallo, M., Polymeropoulos, M. H. & Pastan, I. (1996) *Genome Res.* **6**, 187–194.
38. Bischoff, J. R., Anderson, L., Zhu, Y., Mossie, K., Ng, L., Souza, B., Schryver, B., Flanagan, P., Clairvoyant, F., Ginther, C., *et al.* (1998) *EMBO J.* **17**, 3052–3065.
39. Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B. & Kinzler, K. W. (1997) *Science* **275**, 1787–1790.
40. Lu, L. H. & Gillett, N. (1994) *Cell Vision* **1**, 169–176.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	Ashkenazi et al.
App. No.	:	09/903,925
Filed	:	July 11, 2001
For	:	SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
Examiner	:	Hamud, Fozia M

Group Art Unit 1647

CERTIFICATE OF EXPRESS MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to Commissioner of Patents, Washington D.C. 20231 on:

_____
(Date)

_____

Commissioner of Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION OF AVI ASHKENAZI, Ph.D UNDER 37 C.F.R. § 1.132

I, Avi Ashkenazi, Ph.D. declare and say as follows: -

1. I am Director and Staff Scientist at the Molecular Oncology Department of Genentech, Inc., South San Francisco, CA 94080.

2. I joined Genentech in 1988 as a postdoctoral fellow. Since then, I have investigated a variety of cellular signal transduction mechanisms, including apoptosis, and have developed technologies to modulate such mechanisms as a means of therapeutic intervention in cancer and autoimmune disease. I am currently involved in the investigation of a series of secreted proteins over-expressed in tumors, with the aim to identify useful targets for the development of therapeutic antibodies for cancer treatment.

3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).

4. Gene amplification is a process in which chromosomes undergo changes to contain multiple copies of certain genes that normally exist as a single copy, and is an important factor in the pathophysiology of cancer. Amplification of certain genes (e.g., Myc or Her2/Neu)

gives cancer cells a growth or survival advantage relative to normal cells, and might also provide a mechanism of tumor cell resistance to chemotherapy or radiotherapy.

5. If gene amplification results in over-expression of the mRNA and the corresponding gene product, then it identifies that gene product as a promising target for cancer therapy, for example by the therapeutic antibody approach. Even in the absence of over-expression of the gene product, amplification of a cancer marker gene - as detected, for example, by the reverse transcriptase TaqMan® PCR or the fluorescence *in situ* hybridization (FISH) assays -is useful in the diagnosis or classification of cancer, or in predicting or monitoring the efficacy of cancer therapy. An increase in gene copy number can result not only from intrachromosomal changes but also from chromosomal aneuploidy. It is important to understand that detection of gene amplification can be used for cancer diagnosis even if the determination includes measurement of chromosomal aneuploidy. Indeed, as long as a significant difference relative to normal tissue is detected, it is irrelevant if the signal originates from an increase in the number of gene copies per chromosome and/or an abnormal number of chromosomes.

6. I understand that according to the Patent Office, absent data demonstrating that the increased copy number of a gene in certain types of cancer leads to increased expression of its product, gene amplification data are insufficient to provide substantial utility or well established utility for the gene product (the encoded polypeptide), or an antibody specifically binding the encoded polypeptide. However, even when amplification of a cancer marker gene does not result in significant over-expression of the corresponding gene product, this very absence of gene product over-expression still provides significant information for cancer diagnosis and treatment. Thus, if over-expression of the gene product does not parallel gene amplification in certain tumor types but does so in others, then parallel monitoring of gene amplification and gene product over-expression enables more accurate tumor classification and hence better determination of suitable therapy. In addition, absence of over-expression is crucial information for the practicing clinician. If a gene is amplified but the corresponding gene product is not over-expressed, the clinician accordingly will decide not to treat a patient with agents that target that gene product.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so

made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: Avi Ashkenazi
Avi Ashkenazi, Ph.D.

Date: 9/15/03

SV 455281 v1
9/12/03 3:06 PM (39780.7000)

CURRICULUM VITAE

Avi Ashkenazi

July 2003

Personal:

Date of birth:	29 November, 1956
Address:	1456 Tarrytown Street, San Mateo, CA 94402
Phone:	(650) 578-9199 (home); (650) 225-1853 (office)
Fax:	(650) 225-6443 (office)
Email:	aa@gene.com

Education:

1983:	B.S. in Biochemistry, with honors, Hebrew University, Israel
1986:	Ph.D. in Biochemistry, Hebrew University, Israel

Employment:

1983-1986:	Teaching assistant, undergraduate level course in Biochemistry
1985-1986:	Teaching assistant, graduate level course on Signal Transduction
1986 - 1988:	Postdoctoral fellow, Hormone Research Dept., UCSF, and Developmental Biology Dept., Genentech, Inc., with J. Ramachandran
1988 - 1989:	Postdoctoral fellow, Molecular Biology Dept., Genentech, Inc., with D. Capon
1989 - 1993:	Scientist, Molecular Biology Dept., Genentech, Inc.
1994 -1996:	Senior Scientist, Molecular Oncology Dept., Genentech, Inc.
1996-1997:	Senior Scientist and Interim director, Molecular Oncology Dept., Genentech, Inc.
1997-1990:	Senior Scientist and preclinical project team leader, Genentech, Inc.
1999 -2002:	Staff Scientist in Molecular Oncology, Genentech, Inc.
2002-present:	Staff Scientist and Director in Molecular Oncology, Genentech, Inc.

Awards:

1988:	First prize, The Boehringer Ingelheim Award

Editorial:

Editorial Board Member: Current Biology

Associate Editor, Clinical Cancer Research.

Associate Editor, Cancer Biology and Therapy.

Refereed papers:

1. Gertler, A., Ashkenazi, A., and Madar, Z. Binding sites for human growth hormone and ovine and bovine prolactins in the mammary gland and liver of the lactating cow. *Mol. Cell. Endocrinol.* **34**, 51-57 (1984).
2. Gertler, A., Shamay, A., Cohen, N., Ashkenazi, A., Friesen, H., Levanon, A., Gorecki, M., Aviv, H., Hadari, D., and Vogel, T. Inhibition of lactogenic activities of ovine prolactin and human growth hormone (hGH) by a novel form of a modified recombinant hGH. *Endocrinology* **118**, 720-726 (1986).
3. Ashkenazi, A., Madar, Z., and Gertler, A. Partial purification and characterization of bovine mammary gland prolactin receptor. *Mol. Cell. Endocrinol.* **50**, 79-87 (1987).
4. Ashkenazi, A., Pines, M., and Gertler, A. Down-regulation of lactogenic hormone receptors in Nb2 lymphoma cells by cholera toxin. *Biochemistry Internatl.* **14**, 1065-1072 (1987).
5. Ashkenazi, A., Cohen, R., and Gertler, A. Characterization of lactogen receptors in lactogenic hormone-dependent and independent Nb2 lymphoma cell lines. *FEBS Lett.* **210**, 51-55 (1987).
6. Ashkenazi, A., Vogel, T., Barash, I., Hadari, D., Levanon, A., Gorecki, M., and Gertler, A. Comparative study on in vitro and in vivo modulation of lactogenic and somatotropic receptors by native human growth hormone and its modified recombinant analog. *Endocrinology* **121**, 414-419 (1987).
7. Peralta, E., Winslow, J., Peterson, G., Smith, D., Ashkenazi, A., Ramachandran, J., Schimerlik, M., and Capon, D. Primary structure and biochemical properties of an M2 muscarinic receptor. *Science* **236**, 600-605 (1987).
8. Peralta, E. Ashkenazi, A., Winslow, J., Smith, D., Ramachandran, J., and Capon, D. J. Distincnt primary structures, ligand-binding properties and tissue-specific expression of four human muscarinic acetylcholine receptors. *EMBO J.* **6**, 3923-3929 (1987).
9. Ashkenazi, A., Winslow, J., Peralta, E., Peterson, G., Schimerlik, M., Capon, D., and Ramachandran, J. An M2 muscarinic receptor subtype coupled to both adenylyl cyclase and phosphoinositide turnover. *Science* **238**, 672-675 (1987).

10. Pines, M., Ashkenazi, A., Cohen-Chapnik, N., Binder, L., and Gertler, A. Inhibition of the proliferation of Nb2 lymphoma cells by femtomolar concentrations of cholera toxin and partial reversal of the effect by 12-o-tetradecanoyl-phorbol-13-acetate. *J. Cell. Biochem.* **37**, 119-129 (1988).

11. Peralta, E. Ashkenazi, A., Winslow, J. Ramachandran, J., and Capon, D. Differential regulation of PI hydrolysis and adenylyl cyclase by muscarinic receptor subtypes. *Nature* **334**, 434-437 (1988).

12. Ashkenazi., A. Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functionally distinct G proteins couple different receptors to PI hydrolysis in the same cell. *Cell* **56**, 487-493 (1989).

13. Ashkenazi, A., Ramachandran, J., and Capon, D. Acetylcholine analogue stimulates DNA synthesis in brain-derived cells via specific muscarinic acetylcholine receptor subtypes. *Nature* **340**, 146-150 (1989).

14. Lammare, D., Ashkenazi, A., Fleury, S., Smith, D., Sekaly, R., and Capon, D. The MHC-binding and gp120-binding domains of CD4 are distinct and separable. *Science* **245**, 743-745 (1989).

15. Ashkenazi., A., Presta, L., Marsters, S., Camerato, T., Rosenthal, K., Fendly, B., and Capon, D. Mapping the CD4 binding site for human immunodefficiency virus type 1 by alanine-scanning mutagenesis. *Proc. Natl. Acad. Sci. USA.* **87**, 7150-7154 (1990).

16. Chamow, S., Peers, D., Byrn, R., Mulkerrin, M., Harris, R., Wang, W., Bjorkman, P., Capon, D., and Ashkenazi, A. Enzymatic cleavage of a CD4 immunoadhesin generates crystallizable, biologically active Fd-like fragments. *Biochemistry* **29**, 9885-9891 (1990).

17. Ashkenazi, A., Smith, D., Marsters, S., Riddle, L., Gregory, T., Ho, D., and Capon, D. Resistance of primary isolates of human immunodefficiency virus type 1 to soluble CD4 is independent of CD4-rgp120 binding affinity. *Proc. Natl. Acad. Sci. USA.* **88**, 7056-7060 (1991).

18. Ashkenazi, A., Marsters, S., Capon, D., Chamow, S., Figari., I., Pennica, D., Goeddel., D., Palladino, M., and Smith, D. Protection against endotoxic shock by a tumor necrosis factor receptor immunoadhesin. *Proc. Natl. Acad. Sci. USA.* **88**, 10535-10539 (1991).

19. Moore, J., McKeating, J., Huang, Y., Ashkenazi, A., and Ho, D. Virions of primary HIV-1 isolates resistant to sCD4 neutralization differ in sCD4 affinity and glycoprotein gp120 retention from sCD4-sensitive isolates. *J. Virol.* **66**, 235-243 (1992).

20. Jin, H., Oksenberg, D., Ashkenazi, A., Peroutka, S., Duncan, A., Rozmahel., R., Yang, Y., Mengod, G., Palacios, J., and O'Dowd, B. Characterization of the human 5-hydroxytryptamine$_{1B}$ receptor. *J. Biol. Chem.* **267**, 5735-5738 (1992).

21. Marsters, A., Frutkin, A., Simpson, N., Fendly, B. and Ashkenazi, A. Identification of cysteine-rich domains of the type 1 tumor necrosis receptor involved in ligand binding. *J. Biol. Chem.* **267**, 5747-5750 (1992).

22. Chamow, S., Kogan, T., Peers, D., Hastings, R., Byrn, R., and Ashkenazi, A. Conjugation of sCD4 without loss of biological activity via a novel carbohydrate-directed cross-linking reagent. *J. Biol. Chem.* **267**, 15916-15922 (1992).

23. Oksenberg, D., Marsters, A., O'Dowd, B., Jin, H., Havlik, S., Peroutka, S., and Ashkenazi, A. A single amino-acid difference confers major pharmacologic variation between human and rodent 5-HT$_{1B}$ receptors. *Nature* **360**, 161-163 (1992).

24. Haak-Frendscho, M., Marsters, S., Chamow, S., Peers, D., Simpson, N., and Ashkenazi, A. Inhibition of interferon γ by an interferon γ receptor immunoadhesin. *Immunology* **79**, 594-599 (1993).

25. Penica, D., Lam, V., Weber, R., Kohr, W., Basa, L., Spellman, M., Ashkenazi, Shire, S., and Goeddel, D. Biochemical characterization of the extracellular domain of the 75-kd tumor necrosis factor receptor. *Biochemistry* **32**, 3131-3138. (1993).

26. Barfod, L., Zheng, Y., Kuang, W., Hart, M., Evans, T., Cerione, R., and Ashkenazi, A. Cloning and expression of a human CDC42 GTPase Activating Protein reveals a functional SH3-binding domain. *J. Biol. Chem.* **268**, 26059-26062 (1993).

27. Chamow, S., Zhang, D., Tan, X., Mhtre, S., Marsters, S., Peers, D., Byrn, R., Ashkenazi, A., and Yunghans, R. A humanized bispecific immunoadhesin-antibody that retargets CD3+ effectors to kill HIV-1-infected cells. *J. Immunol.* **153**, 4268-4280 (1994).

28. Means, R., Krantz, S., Luna, J., Marsters, S., and Ashkenazi, A. Inhibition of murine erythroid colony formation in vitro by iterferon γ and correction by interferon γ receptor immunoadhesin. *Blood* **83**, 911-915 (1994).

29. Haak-Frendscho, M., Marsters, S., Mordenti, J., Gillet, N., Chen, S., andAshkenazi, A. Inhibition of TNF by a TNF receptor immunoadhesin: comparison with an anti-TNF mAb. *J. Immunol.* **152**, 1347-1353 (1994).

30. Chamow, S., Kogan, T., Venuti, M., Gadek, T., Peers, D., Mordenti, J., Shak, S., and Ashkenazi, A. Modification of CD4 immunoadhesin with monomethoxy-PEG aldehyde via reductive alkilation. *Bioconj. Chem.* **5**, 133-140 (1994).

31. Jin, H., Yang, R., Marsters, S., Bunting, S., Wurm, F., Chamow, S., and Ashkenazi, A. Protection against rat endotoxic shock by p55 tumor necrosis factor (TNF) receptor immunoadhesin: comparison to anti-TNF monoclonal antibody. *J. Infect. Diseases* **170**, 1323-1326 (1994).

32. Beck, J., Marsters, S., Harris, R., Ashkenazi, A., and Chamow, S. Generation of soluble interleukin-1 receptor from an immunoadhesin by specific cleavage. *Mol. Immunol.* **31**, 1335-1344 (1994).

33. Pitti, B., Marsters, M., Haak-Frendscho, M., Osaka, G., Mordenti, J., Chamow, S., and Ashkenazi, A. Molecular and biological properties of an interleukin-1 receptor immunoadhesin. *Mol. Immunol.* **31**, 1345-1351 (1994).

34. Oksenberg, D., Havlik, S., Peroutka, S., and Ashkenazi, A. The third intracellular loop of the 5-HT2 receptor specifies effector coupling. *J. Neurochem.* **64**, 1440-1447 (1995).

35. Bach, E., Szabo, S., Dighe, A., Ashkenazi, A., Aguet, M., Murphy, K., and Schreiber, R. Ligand-induced autoregulation of IFN-γ receptor β chain expression in T helper cell subsets. *Science* **270**, 1215-1218 (1995).

36. Jin, H., Yang, R., Marsters, S., Ashkenazi, A., Bunting, S., Marra, M., Scott, R., and Baker, J. Protection against endotoxic shock by bactericidal/permeability-increasing protein in rats. *J. Clin. Invest.* **95**, 1947-1952 (1995).

37. Marsters, S., Penica, D., Bach, E., Schreiber, R., and Ashkenazi, A. Interferon γ signals via a high-affinity multisubunit receptor complex that contains two types of polypeptide chain. *Proc. Natl. Acad. Sci. USA.* **92**, 5401-5405 (1995).

38. Van Zee, K., Moldawer, L., Oldenburg, H., Thompson, W., Stackpole, S., Montegut, W., Rogy, M., Meschter, C., Gallati, H., Schiller, C., Richter, W., Loetcher, H., Ashkenazi, A., Chamow, S., Wurm, F., Calvano, S., Lowry, S., and Lesslauer, W. Protection against lethal E. coli bacteremia in baboons by pretreatment with a 55-kDa TNF receptor-Ig fusion protein, Ro45-2081. *J. Immunol.* **156**, 2221-2230 (1996).

39. Pitti, R., Marsters, S., Ruppert, S., Donahue, C., Moore, A., and Ashkenazi, A. Induction of apoptosis by Apo-2 Ligand, a new member of the tumor necrosis factor cytokine family. *J. Biol. Chem.* **271**, 12687-12690 (1996).

40. Marsters, S., Pitti, R., Donahue, C., Rupert, S., Bauer, K., and Ashkenazi, A. Activation of apoptosis by Apo-2 ligand is independent of FADD but blocked by CrmA. *Curr. Biol.* **6**, 1669-1676 (1996).

41. Marsters, S., Skubatch, M., Gray, C., and Ashkenazi, A. Herpesvirus entry mediator, a novel member of the tumor necrosis factor receptor family, activates the NF-κB and AP-1 transcription factors. *J. Biol. Chem.* **272**, 14029-14032 (1997).

42. Sheridan, J., Marsters, S., Pitti, R., Gurney, A., Skubatch, M., Baldwin, D., Ramakrishnan, L., Gray, C., Baker, K., Wood, W.I., Goddard, A., Godowski, P., and Ashkenazi, A. Control of TRAIL-induced apoptosis by a family of signaling and decoy receptors. *Science* **277**, 818-821 (1997).

43. Marsters, S., Sheridan, J., Pitti, R., Gurney, A., Skubatch, M., Balswin, D., Huang, A., Yuan, J., Goddard, A., Godowski, P., and Ashkenazi, A. A novel receptor for Apo2L/TRAIL contains a truncated death domain. *Curr. Biol.* **7**, 1003-1006 (1997).

44. Marsters, A., Sheridan, J., Pitti, R., Brush, J., Goddard, A., and Ashkenazi, A. Identification of a ligand for the death-domain-containing receptor Apo3. *Curr. Biol.* **8**, 525-528 (1998).

45. Rieger, J., Naumann, U., Glaser, T., Ashkenazi, A., and Weller, M. Apo2 ligand: a novel weapon against malignant glioma? *FEBS Lett.* **427**, 124-128 (1998).

46. Pender, S., Fell, J., Chamow, S., Ashkenazi, A., and MacDonald, T. A p55 TNF receptor immunoadhesin prevents T cell mediated intestinal injury by inhibiting matrix metalloproteinase production. *J. Immunol.* **160**, 4098-4103 (1998).

47. Pitti, R., Marsters, S., Lawrence, D., Roy, Kischkel, F., M., Dowd, P., Huang, A., Donahue, C., Sherwood, S., Baldwin, D., Godowski, P., Wood, W., Gurney, A., Hillan, K., Cohen, R., Goddard, A., Botstein, D., and Ashkenazi, A. Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* **396**, 699-703 (1998).

48. Mori, S., Marakami-Mori, K., Nakamura, S., Ashkenazi, A., and Bonavida, B. Sensitization of AIDS Kaposi's sarcoma cells to Apo-2 ligand-induced apoptosis by actinomycin D. *J. Immunol.* **162**, 5616-5623 (1999).

49. Gurney, A. Marsters, S., Huang, A., Pitti, R., Mark, M., Baldwin, D., Gray, A., Dowd, P., Brush, J., Heldens, S., Schow, P., Goddard, A., Wood, W., Baker, K., Godowski, P., and Ashkenazi, A. Identification of a new member of the tumor necrosis factor family and its receptor, a human ortholog of mouse GITR. *Curr. Biol.* **9**, 215-218 (1999).

50. Ashkenazi, A., Pai, R., Fong, s., Leung, S., Lawrence, D., Marsters, S., Blackie, C., Chang, L., McMurtrey, A., Hebert, A., DeForge, L., Khoumenis, I., Lewis, D., Harris, L., Bussiere, J., Koeppen, H., Shahrokh, Z., and Schwall, R. Safety and anti-tumor activity of recombinant soluble Apo2 ligand. *J. Clin. Invest.* **104**, 155-162 (1999).

51. Chuntharapai, A., Gibbs, V., Lu, J., Ow, A., Marsters, S., Ashkenazi, A., De Vos, A., Kim, K.J. Determination of residues involved in ligand binding and signal transmissiion in the human IFN-α receptor 2. *J. Immunol.* **163**, 766-773 (1999).

52. Johnsen, A.-C., Haux, J., Steinkjer, B., Nonstad, U., Egeberg, K., Sundan, A., Ashkenazi, A., and Espevik, T. Regulation of Apo2L/TRAIL expression in NK cells – involvement in NK cell-mediated cytotoxicity. *Cytokine* **11**, 664-672 (1999).

53. Roth, W., Isenmann, S., Naumann, U., Kugler, S., Bahr, M., Dichgans, J., Ashkenazi, A., and Weller, M. Eradication of intracranial human malignant glioma xenografts by Apo2L/TRAIL. *Biochem. Biophys. Res. Commun.* **265**, 479-483 (1999).

54. Hymowitz, S.G., Christinger, H.W., Fuh, G., Ultsch, M., O'Connell, M., Kelley, R.F., Ashkenazi, A. and de Vos, A.M. Triggering Cell Death: The Crystal Structure of Apo2L/TRAIL in a Complex with Death Receptor 5. *Molec. Cell* **4**, 563–571 (1999).

55. Hymowitz, S.G., O'Connel, M.P., Utsch, M.H., Hurst, A., Totpal, K., Ashkenazi, A., de Vos, A.M., Kelley, R.F. A unique zinc-binding site revealed by a high-resolution X-ray structure of homotrimeric Apo2L/TRAIL. *Biochemistry* **39**, 633-640 (2000).

56. Zhou, Q., Fukushima, P., DeGraff, W., Mitchell, J.B., Stetler-Stevenson, M., Ashkenazi, A., and Steeg, P.S. Radiation and the Apo2L/TRAIL apoptotic pathway preferentially inhibit the colonization of premalignant human breast cancer cells overexpressing cyclin D1. *Cancer Res.* **60**, 2611-2615 (2000).

57. Kischkel, F.C., Lawrence, D. A., Chuntharapai, A., Schow, P., Kim, J., and Ashkenazi, A. Apo2L/TRAIL-dependent recruitment of endogenous FADD and Caspase-8 to death receptors 4 and 5. *Immunity* **12**, 611-620 (2000).

58. Yan, M., Marsters, S.A., Grewal, I.S., Wang, H., *Ashkenazi, A., and *Dixit, V.M. Identification of a receptor for BlyS demonstrates a crucial role in humoral immunity. *Nature Immunol.* **1**, 37-41 (2000).

59. Marsters, S.A., Yan, M., Pitti, R.M., Haas, P.E., Dixit, V.M., and Ashkenazi, A. Interaction of the TNF homologues BLyS and APRIL with the TNF receptor homologues BCMA and TACI. *Curr. Biol.* **10**, 785-788 (2000).

60. Kischkel, F.C., and Ashkenazi, A. Combining enhanced metabolic labeling with immunoblotting to detect interactions of endogenous cellular proteins. *Biotechniques* **29**, 506-512 (2000).

61. Lawrence, D., Shahrokh, Z., Marsters, S., Achilles, K., Shih, D. Mounho, B., Hillan, K., Totpal, K. DeForge, L., Schow, P., Hooley, J., Sherwood, S., Pai, R., Leung, S., Khan, L., Gliniak, B., Bussiere, J., Smith, C., Strom, S., Kelley, S., Fox, J., Thomas, D., and Ashkenazi, A. Differential hepatocyte toxicity of recombinant Apo2L/TRAIL versions. *Nature Med.* **7**, 383-385 (2001).

62. Chuntharapai, A., Dodge, K., Grimmer, K., Schroeder, K., Martsters, S.A., Koeppen, H., Ashkenazi, A., and Kim, K.J. Isotype-dependent inhibition of tumor growth in vivo by monoclonal antibodies to death receptor 4. *J. Immunol.* **166**, 4891-4898 (2001).

63. Pollack, I.F., Erff, M., and Ashkenazi, A. Direct stimulation of apoptotic signaling by soluble Apo2L/tumor necrosis factor-related apoptosis-inducing ligand leads to selective killing of glioma cells. *Clin. Cancer Res.* **7**, 1362-1369 (2001).

64. Wang, H., Marsters, S.A., Baker, T., Chan, B., Lee, W.P., Fu, L., Tumas, D., Yan, M., Dixit, V.M., *Ashkenazi, A., and *Grewal, I.S. TACI-ligand interactions are required for T cell activation and collagen-induced arthritis in mice. *Nature Immunol.* 2, 632-637 (2001).

65. Kischkel, F.C., Lawrence, D. A., Tinel, A., Virmani, A., Schow, P., Gazdar, A., Blenis, J., Arnott, D., and Ashkenazi, A. Death receptor recruitment of endogenous caspase-10 and apoptosis initiation in the absence of caspase-8. *J. Biol. Chem.* 276, 46639-46646 (2001).

66. LeBlanc, H., Lawrence, D.A., Varfolomeev, E., Totpal, K., Morlan, J., Schow, P., Fong, S., Schwall, R., Sinicropi, D., and Ashkenazi, A Tumor cell resistance to death receptor induced apoptosis through mutational inactivation of the proapoptotitc Bcl-2 homolog Bax. *Nature Med.* 8, 274-281 (2002).

67. Miller, K., Meng, G., Liu, J., Hurst, A., Hsei, V., Wong, W-L., Ekert, R., Lawrence, D., Sherwood, S., DeForge, L., Gaudreault., Keller, G., Sliwkowski, M., Ashkenazi, A., and Presta, L. Design, Construction, and analyses of multivalent antibodies. *J. Immunol.* 170, 4854-4861 (2003).

68. Varfolomeev, E., Kischkel, F., Martin, F., Wanh, H., Lawrence, D., Olsson, C., Tom, L., Erickson, S., French, D., Schow, P., Grewal, I. and Ashkenazi, A. Immune system development in APRIL knockout mice. Submitted.

Review articles:

1. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D., J. Functional role of muscarinic acetylcholine receptor subtype diversity. *Cold Spring Harbor Symposium on Quantitative Biology*. **LIII**, 263-272 (1988).
2. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functional diversity of muscarinic receptor subtypes in cellular signal transduction and growth. *Trends Pharmacol. Sci.* Dec Supplement, 12-21 (1989).
3. Chamow, S., Duliege, A., Ammann, A., Kahn, J., Allen, D., Eichberg, J., Byrn, R., Capon, D., Ward, R., and Ashkenazi, A. CD4 immunoadhesins in anti-HIV therapy: new developments. *Int. J. Cancer* Supplement 7, 69-72 (1992).
4. Ashkenazi, A., Capon, and D. Ward, R. Immunoadhesins. *Int. Rev. Immunol.* **10**, 217-225 (1993).
5. Ashkenazi, A., and Peralta, E. Muscarinic Receptors. In *Handbook of Receptors and Channels*. (S. Peroutka, ed.), CRC Press, Boca Raton, Vol. **I**, p. 1-27, (1994).
6. Krantz, S. B., Means, R. T., Jr., Lina, J., Marsters, S. A., and Ashkenazi, A. Inhibition of erythroid colony formation in vitro by gamma interferon. In *Molecular Biology of Hematopoiesis* (N. Abraham, R. Shadduck, A. Levine F. Takaku, eds.) Intercept Ltd. Paris, Vol. **3**, p. 135-147 (1994).
7. Ashkenazi, A. Cytokine neutralization as a potential therapeutic approach for SIRS and shock. *J. Biotechnology in Healthcare* **1**, 197-206 (1994).
8. Ashkenazi, A., and Chamow, S. M. Immunoadhesins: an alternative to human monoclonal antibodies. *Immunomethods: A companion to Methods in Enzimology* **8**, 104-115 (1995).
9. Chamow, S., and Ashkenazi, A. Immunoadhesins: Principles and Applications. *Trends Biotech.* **14**, 52-60 (1996).
10. Ashkenazi, A., and Chamow, S. M. Immunoadhesins as research tools and therapeutic agents. *Curr. Opin. Immunol.* **9**, 195-200 (1997).
11. Ashkenazi, A., and Dixit, V. Death receptors: signaling and modulation. *Science* **281**, 1305-1308 (1998).
12. Ashkenazi, A., and Dixit, V. Apoptosis control by death and decoy receptors. *Curr. Opin. Cell. Biol.* **11**, 255-260 (1999).

13. Ashkenazi, A. Chapters on Apo2L/TRAIL; DR4, DR5, DcR1, DcR2; and DcR3. Online Cytokine Handbook (www.apnet.com/cytokinereference/).
14. Ashkenazi, A. Targeting death and decoy receptors of the tumor necrosis factor superfamily. *Nature Rev. Cancer* **2**, 420-430 (2002).
15. LeBlanc, H. and Ashkenazi, A. Apoptosis signaling by Apo2L/TRAIL. *Cell Death and Differentiation* 10, 66-75 (2003).
16. Almasan, A. and Ashkenazi, A. Apo2L/TRAIL: apoptosis signaling, biology, and potential for cancer therapy. Cytokine and Growth Factor Reviews 14, 337-348 (2003).

Book:

Antibody Fusion Proteins (Chamow, S., and Ashkenazi, A., eds., John Wiley and Sons Inc.) (1999).

Talks:

1. Resistance of primary HIV isolates to CD4 is independent of CD4-gp120 binding affinity. UCSD Symposium, HIV Disease: Pathogenesis and Therapy. Greenelefe, FL, March 1991.
2. Use of immuno-hybrids to extend the half-life of receptors. IBC conference on Biopharmaceutical Halflife Extension. New Orleans, LA, June 1992.
3. Results with TNF receptor Immunoadhesins for the Treatment of Sepsis. IBC conference on Endotoxemia and Sepsis. Philadelphia, PA, June 1992.
4. Immunoadhesins: an alternative to human antibodies. IBC conference on Antibody Engineering. San Diego, CA, December 1993.
5. Tumor necrosis factor receptor: a potential therapeutic for human septic shock. American Society for Microbiology Meeting, Atlanta, GA, May 1993.
6. Protective efficiacy of TNF receptor immunoadhesin vs anti-TNF monoclonal antibody in a rat model for endotoxic shock. 5th International Congress on TNF. Asilomar, CA, May 1994.
7. Interferon-γ signals via a multisubunit receptor complex that contains two types of polypeptide chain. American Association of Immunologists Conference. San Franciso, CA, July 1995.
8. Immunoadhesins: Principles and Applications. Gordon Research Conference on Drug Delivery in Biology and Medicine. Ventura, CA, February 1996.

9. Apo-2 Ligand, a new member of the TNF family that induces apoptosis in tumor cells. Cambridge Symposium on TNF and Related Cytokines in Treatment of Cancer. Hilton-Head, NC, March 1996.
10. Induction of apoptosis by Apo2 Ligand. American Society for Biochemistry and Molecular Biology, Symposium on Growth Factors and Cytokine Receptors. New Orleans, LA, June, 1996.
11. Apo2 ligand, an extracellular trigger of apoptosis. 2nd Clontech Symposium, Palo Alto, CA, October 1996.
12. Regulation of apoptosis by members of the TNF ligand and receptor families. Stanford University School of Medicine, Palo Alto, CA, December 1996.
13. Apo-3: anovel receptor that regulates cell death and inflammation. 4th International Congress on Immune Consequences of Trauma, Shock, and Sepsis. Munich, Germany, March 1997.
14. New members of the TNF ligand and receptor families that regulate apoptosis, inflammation, and immunity. UCLA School of Medicine, LA, CA, March 1997.
15. Immunoadhesins: an alternative to monoclonal antibodies. 5th World Conference on Bispecific Antibodies. Volendam, Holland, June 1997.
16. Control of Apo2L signaling. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. Cold Spring Harbor, New York. September, 1997.
17. Chairman and speaker, Apoptosis Signaling session. IBC's 4th Annual Conference on Apoptosis. San Diego, CA., October 1997.
18. Control of Apo2L signaling by death and decoy receptors. American Association for the Advancement of Science. Philladelphia, PA, February 1998.
19. Apo2 ligand and its receptors. American Society of Immunologists. San Francisco, CA, April 1998.
20. Death receptors and ligands. 7th International TNF Congress. Cape Cod, MA, May 1998.
21. Apo2L as a potential therapeutic for cancer. UCLA School of Medicine. LA, CA, June 1998.
22. Apo2L as a potential therapeutic for cancer. Gordon Research Conference on Cancer Chemotherapy. New London, NH, July 1998.
23. Control of apoptosis by Apo2L. Endocrine Society Conference, Stevenson, WA, August 1998.
24. Control of apoptosis by Apo2L. International Cytokine Society Conference, Jerusalem, Israel, October 1998.

25. Apoptosis control by death and decoy receptors. American Association for Cancer Research Conference, Whistler, BC, Canada, March 1999.
26. Apoptosis control by death and decoy receptors. American Society for Biochemistry and Molecular Biology Conference, San Francisco, CA, May 1999.
27. Apoptosis control by death and decoy receptors. Gordon Research Conference on Apoptosis, New London, NH, June 1999.
28. Apoptosis control by death and decoy receptors. Arthritis Foundation Research Conference, Alexandria GA, Aug 1999.
29. Safety and anti-tumor activity of recombinant soluble Apo2L/TRAIL. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. . Cold Spring Harbor, NY, September 1999.
30. The Apo2L/TRAIL system: therapeutic potential. American Association for Cancer Research, Lake Tahoe, NV, Feb 2000.
31. Apoptosis and cancer therapy. Stanford University School of Medicine, Stanford, CA, Mar 2000.
32. Apoptosis and cancer therapy. University of Pennsylvania School of Medicine, Philladelphia, PA, Apr 2000.
33. Apoptosis signaling by Apo2L/TRAIL. International Congress on TNF. Trondheim, Norway, May 2000.
34. The Apo2L/TRAIL system: therapeutic potential. Cap-CURE summit meeting. Santa Monica, CA, June 2000.
35. The Apo2L/TRAIL system: therapeutic potential. MD Anderson Cancer Center. Houston, TX, June 2000.
36. Apoptosis signaling by Apo2L/TRAIL. The Protein Society, 14th Symposium. San Diego, CA, August 2000.
37. Anti-tumor activity of Apo2L/TRAIL. AAPS annual meeting. Indianapolis, IN Aug 2000.
38. Apoptosis signaling and anti-cancer potential of Apo2L/TRAIL. Cancer Research Institute, UC San Francisco, CA, September 2000.
39. Apoptosis signaling by Apo2L/TRAIL. Kenote address, TNF family Minisymposium, NIH. Bethesda, MD, September 2000.
40. Death receptors: signaling and modulation. Keystone symposium on the Molecular basis of cancer. Taos, NM, Jan 2001.
41. Preclinical studies of Apo2L/TRAIL in cancer. Symposium on Targeted therapies in the treatment of lung cancer. Aspen, CO, Jan 2001.

42. Apoptosis signaling by Apo2L/TRAIL. Wiezmann Institute of Science, Rehovot, Israel, March 2001.
43. Apo2L/TRAIL: Apoptosis signaling and potential for cancer therapy. Weizmann Institute of Science, Rehovot, Israel, March 2001.
44. Targeting death receptors in cancer with Apo2L/TRAIL. Cell Death and Disease conference, North Falmouth, MA, Jun 2001.
45. Targeting death receptors in cancer with Apo2L/TRAIL. Biotechnology Organization conference, San Diego, CA, Jun 2001.
46. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Gordon Research Conference on Apoptosis, Oxford, UK, July 2001.
47. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Cleveland Clinic Foundation, Cleveland, OH, Oct 2001.
48. Apoptosis signaling by death receptors: overview. International Society for Interferon and Cytokine Research conference, Cleveland, OH, Oct 2001.
49. Apoptosis signaling by death receptors. American Society of Nephrology Conference. San Francisco, CA, Oct 2001.
50. Targeting death receptors in cancer. Apoptosis: commercial opportunities. San Diego, CA, Apr 2002.
51. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Kimmel Cancer Research Center, Johns Hopkins University, Baltimore MD. May 2002.
52. Apoptosis control by Apo2L/TRAIL. (Keynote Address) University of Alabama Cancer Center Retreat, Birmingham, Ab. October 2002.
53. Apoptosis signaling by Apo2L/TRAIL. (Session co-chair) TNF international conference. San Diego, CA. October 2002.
54. Apoptosis signaling by Apo2L/TRAIL. Swiss Institute for Cancer Research (ISREC). Lausanne, Swizerland. Jan 2003.
55. Apoptosis induction with Apo2L/TRAIL. Conference on New Targets and Innovative Strategies in Cancer Treatment. Monte Carlo. February 2003.
56. Apoptosis signaling by Apo2L/TRAIL. Hermelin Brain Tumor Center Symposium on Apoptosis. Detroit, MI. April 2003.
57. Targeting apoptosis through death receptors. Sixth Annual Conference on Targeted Therapies in the Treatment of Breast Cancer. Kona, Hawaii. July 2003.
58. Targeting apoptosis through death receptors. Second International Conference on Targeted Cancer Therapy. Washington, DC. Aug 2003.

Issued Patents:

1. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,329,028 (Jul 12, 1994).
2. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,605,791 (Feb 25, 1997).
3. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,889,155 (Jul 27, 1999).
4. Ashkenazi, A., APO-2 Ligand. US patent 6,030,945 (Feb 29, 2000).
5. Ashkenazi, A., Chuntharapai, A., Kim, J., APO-2 ligand antibodies. US patent 6, 046, 048 (Apr 4, 2000).
6. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,124,435 (Sep 26, 2000).
7. Ashkenazi, A., Chuntharapai, A., Kim, J., Method for making monoclonal and cross-reactive antibodies. US patent 6,252,050 (Jun 26, 2001).
8. Ashkenazi, A. APO-2 Receptor. US patent 6,342,369 (Jan 29, 2002).
9. Ashkenazi, A. Fong, S., Goddard, A., Gurney, A., Napier, M., Tumas, D., Wood, W. A-33 polypeptides. US patent 6,410,708 (Jun 25, 2002).
10. Ashkenazi, A. APO-3 Receptor. US patent 6,462,176 B1 (Oct 8, 2002).
11. Ashkenazi, A. APO-2LI and APO-3 polypeptide antibodies. US patent 6,469,144 B1 (Oct 22, 2002).
12. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,582,928B1 (Jun 24, 2003).

DECLARATION OF PAUL POLAKIS, Ph.D.

I, Paul Polakis, Ph.D., declare and say as follows:

1. I was awarded a Ph.D. by the Department of Biochemistry of the Michigan State University in 1984. My scientific Curriculum Vitae is attached to and forms part of this Declaration (Exhibit A).

2. I am currently employed by Genentech, Inc. where my job title is Staff Scientist. Since joining Genentech in 1999, one of my primary responsibilities has been leading Genentech's Tumor Antigen Project, which is a large research project with a primary focus on identifying tumor cell markers that find use as targets for both the diagnosis and treatment of cancer in humans.

3. As part of the Tumor Antigen Project, my laboratory has been analyzing differential expression of various genes in tumor cells relative to normal cells. The purpose of this research is to identify proteins that are abundantly expressed on certain tumor cells and that are either (i) not expressed, or (ii) expressed at lower levels, on corresponding normal cells. We call such differentially expressed proteins "tumor antigen proteins". When such a tumor antigen protein is identified, one can produce an antibody that recognizes and binds to that protein. Such an antibody finds use in the diagnosis of human cancer and may ultimately serve as an effective therapeutic in the treatment of human cancer.

4. In the course of the research conducted by Genentech's Tumor Antigen Project, we have employed a variety of scientific techniques for detecting and studying differential gene expression in human tumor cells relative to normal cells, at genomic DNA, mRNA and protein levels. An important example of one such technique is the well known and widely used technique of microarray analysis which has proven to be extremely useful for the identification of mRNA molecules that are differentially expressed in one tissue or cell type relative to another. In the course of our research using microarray analysis, we have identified approximately 200 gene transcripts that are present in human tumor cells at significantly higher levels than in corresponding normal human cells. To date, we have generated antibodies that bind to about 30 of the tumor antigen proteins expressed from these differentially expressed gene transcripts and have used these antibodies to quantitatively determine the level of production of these tumor antigen proteins in both human cancer cells and corresponding normal cells. We have then compared the levels of mRNA and protein in both the tumor and normal cells analyzed.

5. From the mRNA and protein expression analyses described in paragraph 4 above, we have observed that there is a strong correlation between changes in the level of mRNA present in any particular cell type and the level of protein

expressed from that mRNA in that cell type. In approximately 80% of our observations we have found that increases in the level of a particular mRNA correlates with changes in the level of protein expressed from that mRNA when human tumor cells are compared with their corresponding normal cells.

6. Based upon my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4 and 5 above and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell. In fact, it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein. While there have been published reports of genes for which such a correlation does not exist, it is my opinion that such reports are exceptions to the commonly understood general rule that increased mRNA levels are predictive of corresponding increased levels of the encoded protein.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 5/07/04

By: Paul Polakis

Paul Polakis, Ph.D.

SV 2031808 v1

CURRICULUM VITAE

PAUL G. POLAKIS
Staff Scientist
Genentech, Inc
1 DNA Way, MS#40
S. San Francisco, CA 94080

EDUCATION:

Ph.D., Biochemistry, Department of Biochemistry,
Michigan State University (1984)

B.S., Biology. College of Natural Science, Michigan State University (1977)

PROFESSIONAL EXPERIENCE:

2002-present	Staff Scientist, Genentech, Inc S. San Francisco, CA
1999- 2002	Senior Scientist, Genentech, Inc., S. San Francisco, CA
1997 -1999	Research Director Onyx Pharmaceuticals, Richmond, CA
1992- 1996	Senior Scientist, Project Leader, Onyx Pharmaceuticals, Richmond, CA
1991-1992	Senior Scientist, Chiron Corporation, Emeryville, CA.
1989-1991	Scientist, Cetus Corporation, Emeryville CA.
1987-1989	Postdoctoral Research Associate, Genentech, Inc., South SanFrancisco, CA.
1985-1987	Postdoctoral Research Associate, Department of Medicine, Duke University Medical Center, Durham, NC

1984-1985	Assistant Professor, Department of Chemistry, Oberlin College, Oberlin, Ohio
1980-1984	Graduate Research Assistant, Department of Biochemistry, Michigan State University East Lansing, Michigan

PUBLICATIONS:

1. **Polakis, P G.** and Wilson, J. E. 1982 Purification of a Highly Bindable Rat Brain Hexokinase by High Performance Liquid Chromatography. **Biochem. Biophys. Res. Commun**. 107, 937-943.

2. **Polakis, P.G.** and Wilson, J. E. 1984 Proteolytic Dissection of Rat Brain Hexokinase: Determination of the Cleavage Pattern during Limited Digestion with Trypsin. **Arch. Biochem. Biophys**. 234, 341-352.

3. **Polakis, P. G.** and Wilson, J. E. 1985 An Intact Hydrophobic N-Terminal Sequence is Required for the Binding Rat Brain Hexokinase to Mitochondria. **Arch. Biochem. Biophys**. 236, 328-337.

4. Uhing, R.J., **Polakis,P.G.** and Snyderman, R. 1987 Isolaton of GTP-binding Proteins from Myeloid HL60 Cells. **J. Biol. Chem**. 262, 15575-15579.

5. **Polakis, P.G.**, Uhing, R.J. and Snyderman, R. 1988 The Formylpeptide Chemoattractant Receptor Copurifies with a GTP-binding Protein Containing a Distinct 40 kDa Pertussis Toxin Substrate. **J. Biol. Chem**. 263, 4969-4979.

6. Uhing, R. J., Dillon, S., **Polakis, P. G.**, Truett, A. P. and Snyderman, R. 1988 Chemoattractant Receptors and Signal Transduction Processes in Cellular and Molecular Aspects of Inflammation (Poste, G. and Crooke, S. T. eds.) pp 335-379.

7. **Polakis, P.G.**, Evans, T. and Snyderman 1989 Multiple Chromatographic Forms of the Formylpeptide Chemoattractant Receptor and their Relationship to GTP-binding Proteins. **Biochem. Biophys. Res. Commun**. 161, 276-283.

8. **Polakis, P. G.**, Snyderman, R. and Evans, T. 1989 Characterization of G25K, a GTP-binding Protein Containing a Novel Putative Nucleotide Binding Domain. **Biochem. Biophys. Res. Comun.** 160, 25-32.

9. **Polakis, P.**, Weber,R.F., Nevins,B., Didsbury, J. Evans,T. and Snyderman, R. 1989 Identification of the ral and rac1 Gene Products, Low Molecular Mass GTP-binding Proteins from Human Platelets. **J. Biol. Chem**. 264, 16383-16389.

10. Snyderman, R., Perianin, A., Evans, T., **Polakis, P.** and Didsbury, J. 1989 G Proteins and Neutrophil Function. In ADP-Ribosylating Toxins and G Proteins: Insights into Signal Transduction. (J. Moss and M. Vaughn, eds.) Amer. Soc. Microbiol. pp. 295-323.

11. Hart, M.J., **Polakis, P.G.**, Evans, T. and Cerrione, R.A. 1990 The Identification and Charaterization of an Epidermal Growth Factor-Stimulated Phosphorylation of a Specific Low Molecular Mass GTP-binding Protein in a Reconstituted Phospholipid Vesicle System. **J. Biol. Chem**. 265, 5990-6001.

12. Yatani, A., Okabe, K., **Polakis, P.** Halenbeck, R. McCormick, F. and Brown, A. M. 1990 ras p21 and GAP Inhibit Coupling of Muscarinic Receptors to Atrial K^+ Channels. **Cell**. 61, 769-776.

13. Munemitsu, S., Innis, M.A., Clark, R., McCormick, F., Ullrich, A. and **Polakis, P.G.** 1990 Molecular Cloning and Expression of a G25K cDNA, the Human Homolog of the Yeast Cell Cycle Gene CDC42. **Mol. Cell. Biol**. 10, 5977-5982.

14. Polakis, P.G. Rubinfeld, B. Evans, T. and McCormick, F. 1991 Purification of Plasma Membrane-Associated GTPase Activating Protein (GAP) Specific for rap-1/krev-1 from HL60 Cells. **Proc. Natl. Acad. Sci. USA** 88, 239-243.

15. Moran, M. F., **Polakis, P.**, McCormick, F., Pawson, T. and Ellis, C. 1991 Protein Tyrosine Kinases Regulate the Phosphorylation, Protein Interactions, Subcellular Distribution, and Activity of p21ras GTPase Activating Protein. **Mol. Cell. Biol**. 11, 1804-1812

16. Rubinfeld, B., Wong, G., Bekesi, E. Wood, A. McCormick, F. and **Polakis, P. G.** 1991 A Synthetic Peptide Corresponding to a Sequence in the GTPase Activating Protein Inhibits $p21^{ras}$ Stimulation and Promotes Guanine Nucleotide Exchange. **Internatl. J. Peptide and Prot. Res**. 38, 47-53.

17. Rubinfeld, B., Munemitsu, S., Clark, R., Conroy, L., Watt, K., Crosier, W., McCormick, F., and **Polakis, P.** 1991 Molecular Cloning of a GTPase Activating Protein Specific for the Krev-1 Protein $p21^{rap1}$. **Cell** 65, 1033-1042.

18. Zhang, K. Papageorge, A., G., Martin, P., Vass, W. C., Olah, Z., **Polakis, P.**, McCormick, F. and Lowy, D, R. 1991 Heterogenous Amino Acids in RAS and Rap1A Specifying Sensitivity to GAP Proteins. **Science** 254, 1630-1634.

19. Martin, G., Yatani, A., Clark, R., **Polakis, P.**, Brown, A. M. and McCormick, F. 1992 GAP Domains Responsible for $p21^{ras}$-dependent Inhibition of Muscarinic Atrial K^+ Channel Currents. **Science** 255, 192-194.

20. McCormick, F., Martin, G. A., Clark, R., Bollag, G. and **Polakis, P** . 1992 Regulation of p21ras by GTPase Activating Proteins. Cold Spring Harbor **Symposia on Quantitative Biology**. Vol. 56, 237-241.

21. Pronk, G. B., **Polakis, P.**, Wong, G., deVries-Smits, A. M., Bos J. L. and McCormick, F. 1992 $p60^{v\text{-}src}$ Can Associate with and Phosphorylate the $p21^{ras}$ GTPase Activating Protein. **Oncogene** 7,389-394.

22. Polakis P. and McCormick, F. 1992 Interactions Between $p21^{ras}$ Proteins and Their GTPase Activating Proteins. In **Cancer Surveys** (Franks, L. M., ed.) 12, 25-42.

23. Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M., **Polakis,P.** and McCormick, F. 1992 Molecular coloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. **Cell** 69, 551-558.

24. Polakis, P., Rubinfeld, B. and McCormick, F. 1992 Phosphorylation of rap1GAP in vivo and by cAMP-dependent Kinase and the Cell Cycle $p34^{cdc2}$ Kinase in vitro. **J. Biol. Chem.** 267, 10780-10785.

25. McCabe, P.C., Haubrauck, H., **Polakis, P.**, McCormick, F., and Innis, M. A. 1992 Functional Interactions Between $p21^{rap1A}$ and Components of the Budding pathway of *Saccharomyces cerevisiae*. **Mol. Cell. Biol.** 12, 4084-4092.

26. Rubinfeld, B., Crosier, W.J., Albert, I., Conroy,L., Clark, R., McCormick, F. and **Polakis, P.** 1992 Localization of the rap1GAP Catalytic Domain and Sites of Phosphorylation by Mutational Analysis. **Mol. Cell . Biol.** 12, 4634-4642.

27. Ando, S., Kaibuchi, K., Sasaki, K., Hiraoka, T., Nishiyama, T., Mizuno, T., Asada, M., Nunoi, H., Matsuda, I., Matsuura, Y., **Polakis, P.**, McCormick, F. and Takai, Y. 1992 Post-translational processing of rac p21s is important both for their interaction with the GDP/GTP exchange proteins and for their activation of NADPH oxidase. **J. Biol. Chem.** 267, 25709-25713.

28. Janoueix-Lerosey, I., **Polakis, P.**, Tavitian, A. and deGunzberg, J. 1992 Regulation of the GTPase activity of the ras-related rap2 protein. **Biochem. Biophys. Res. Commun.** 189, 455-464.

29. Polakis, P. 1993 GAPs Specific for the rap1/Krev-1 Protein. in GTP-binding Proteins: the *ras*-superfamily. (J.C. LaCale and F. McCormick, eds.) 445-452.

30. Polakis, P. and McCormick, F. 1993 Structural requirements for the interaction of $p21^{ras}$ with GAP, exchange factors, and its bological effector target. **J. Biol Chem.** 268, 9157-9160.

31. Rubinfeld, B., Souza, B. Albert, I., Muller, O., Chamberlain, S., Masiarz, F., Munemitsu, S. and **Polakis, P.** 1993 Association of the APC gene product with beta- catenin. **Science** 262, 1731-1734.

32. Weiss, J., Rubinfeld, B., **Polakis, P.**, McCormick, F. Cavenee, W. A. and Arden, K. 1993 The gene for human rap1-GTPase activating protein (rap1GAP) maps to chromosome 1p35-1p36.1. **Cytogenet. Cell Genet.** 66, 18-21.

33. Sato, K. Y., **Polakis, P.**, Haubruck, H., Fasching, C. L., McCormick, F. and Stanbridge, E. J. 1994 Analysis of the tumor suppressor acitvity of the K-rev gene in human tumor cell lines. **Cancer Res.** 54, 552-559.

34. Janoueix-Lerosey, I., Fontenay, M., Tobelem, G., Tavitian, A., **Polakis, P.** and DeGunzburg, J. 1994 Phosphorylation of rap1GAP during the cell cycle. **Biochem. Biophys. Res. Commun.** 202, 967-975

35. Munemitsu, S., Souza, B., Mueller, O., Albert, I., Rubinfeld, B., and **Polakis, P.** 1994 The APC gene product associates with microtubules in vivo and affects their assembly in vitro. **Cancer Res.** 54, 3676-3681.

36. Rubinfeld, B. and **Polakis, P.** 1995 Purification of baculovirus produced rap1GAP. **Methods Enz.** 255,31

37. Polakis, P. 1995 Mutations in the APC gene and their implications for protein structure and function. **Current Opinions in Genetics and Development** 5, 66-71

38. Rubinfeld, B., Souza, B., Albert, I., Munemitsu, S. and **Polakis P.** 1995 The APC protein and E-cadherin form similar but independent complexes with α-catenin, β-catenin and Plakoglobin. **J. Biol. Chem.** 270, 5549-5555

39. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B., and **Polakis, P.** 1995 Regulation of intracellular β-catenin levels by the APC tumor suppressor gene. **Proc. Natl. Acad. Sci.** 92, 3046-3050.

40. Lock, P., Fumagalli, S., **Polakis, P.** McCormick, F. and Courtneidge, S. A. 1996 The human p62 cDNA encodes Sam68 and not the rasGAP-associated p62 protein. **Cell** 84, 23-24.

41. Papkoff, J., Rubinfeld, B., Schryver, B. and **Polakis, P.** 1996 Wnt-1 regulates free pools of catenins and stabilizes APC-catenin complexes. **Mol. Cell. Biol.** 16, 2128-2134.

42. Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S. and **Polakis, P.** 1996 Binding of GSK3β to the APC-β-catenin complex and regulation of complex assembly. **Science** 272, 1023-1026.

43. Munemitsu, S., Albert, I., Rubinfeld, B. and **Polakis, P.** 1996 Deletion of amino-terminal structure stabilizes β-catenin in vivo and promotes the hyperphosphorylation of the APC tumor suppressor protein. **Mol. Cell. Biol.**16, 4088-4094.

44. Hart, M. J., Callow, M. G., Sousa, B. and **Polakis P.** 1996 IQGAP1, a calmodulin binding protein witha rasGAP related domain, is a potential effector for cdc42Hs. **EMBO J.** 15, 2997-3005.

45. Nathke, I. S., Adams, C. L., **Polakis, P.**, Sellin, J. and Nelson, W. J. 1996 The adenomatous polyposis coli (APC) tumor suppressor protein is localized to plasma membrane sites involved in active epithelial cell migration. **J. Cell. Biol.** 134, 165-180.

46. Hart, M. J., Sharma, S., elMasry, N., Qui, R-G., McCabe, P., **Polakis, P.** and Bollag, G. 1996 Identification of a novel guanine nucleotide exchange factor for the rho GTPase. **J. Biol. Chem.** 271, 25452.

47. Thomas JE, Smith M, Rubinfeld B, Gutowski M, Beckmann RP, and **Polakis P.** 1996 Subcellular localization and analysis of apparent 180-kDa and 220-kDa proteins of the breast cancer susceptibility gene, BRCA1. **J. Biol. Chem.** 1996 271, 28630-28635

48. Hayashi, S., Rubinfeld, B., Souza, B., **Polakis, P.**, Wieschaus, E., and Levine, A. 1997 A Drosophila homolog of the tumor suppressor adenomatous polyposis coli

down-regulates β -catenin but its zygotic expression is not essential for the regulation of armadillo. **Proc. Natl. Acad. Sci.** 94, 242-247.

49. Vleminckx, K., Rubinfeld, B., **Polakis, P.** and Gumbiner, B. 1997 The APC tumor suppressor protein induces a new axis in Xenopus embryos. **J. Cell. Biol.** 136, 411-420.

50. Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, P. and **Polakis, P.** 1997 Stabilization of β-catenin by genetic defects in melanoma cell lines. **Science** 275, 1790-1792.

51. Polakis, P. The adenomatous polyposis coli (APC) tumor suppressor. 1997 **Biochem. Biophys. Acta**, 1332, F127-F147.

52. Rubinfeld, B., Albert, I., Porfiri, E., Munemitsu, S., and **Polakis, P** 1997 Loss of β-catenin regulation by the APC tumor suppressor protein correlates with loss of structure due to common somatic mutations of the gene. **Cancer Res.** 57, 4624-4630.

53. Porfiri, E., Rubinfeld, B., Albert, I., Hovanes. K., Waterman, M., and **Polakis, P.** 1997 Induction of a β-catenin-LEF-1 complex by wnt-1 and transforming mutants of β-catenin. **Oncogene** 15, 2833-2839.

54. Thomas JE, Smith M, Tonkinson JL, Rubinfeld B, and **Polakis P.,** 1997 Induction of phosphorylation on BRCA1 during the cell cycle and after DNA damage. **Cell Growth Differ**. 8, 801-809.

55. Hart, M., de los Santos, R., Albert, I., Rubinfeld, B., and **Polakis P.**, 1998 Down regulation of β-catenin by human Axin and its association with the adenomatous polyposis coli (APC) tumor suppressor, β-catenin and glycogen synthase kinase 3β. **Current Biology 8,** 573-581.

56. Polakis, P. 1998 The oncogenic activation of β-catenin. **Current Opinions in Genetics and Development** 9, 15-21

57. Matt Hart, Jean-Paul Concordet, Irina Lassot, Iris Albert, Rico del los Santos, Herve Durand, Christine Perret, Bonnee Rubinfled, Florence Margottin, Richard Benarous and **Paul Polakis**. 1999 The F-box protein β–TrCP associates with phosphorylated β–catenin and regulates its activity in the cell. **Current Biology** 9, 207-10.

58. Howard C. Crawford, Barbara M. Fingleton, Bonnee Rubinfeld, **Paul Polakis** and Lynn M. Matrisian 1999 The metalloproteinase matrilysin is a target of β–catenin transactivation in intestinal tumours. **Oncogene** 18, 2883-91.

59. Meng J, Glick JL, **Polakis P**, Casey PJ. 1999 Functional interaction between Galpha(z) and Rap1GAP suggests a novel form of cellular cross-talk. **J Biol Chem.** 17, 36663-9

60. Vijayasurian Easwaran, Virginia Song, **Paul Polakis** and Steve Byers 1999 The ubiquitin-proteosome pathway and serine kinase activity modulate APC mediated regulation of β-catenin-LEF signaling. **J. Biol. Chem**. 274(23):16641-5.

61 **Polakis P**, Hart M and Rubinfeld B. 1999 Defects in the regulation of beta-catenin in colorectal cancer. **Adv Exp Med Biol**. 470, 23-32

62 Shen Z, Batzer A, Koehler JA, **Polakis P**, Schlessinger J, Lydon NB, Moran MF. 1999 Evidence for SH3 domain directed binding and phosphorylation of Sam68 by Src. **Oncogene**. 18, 4647-53

64. Thomas GM, Frame S, Goedert M, Nathke I, **Polakis P**, Cohen P. 1999 A GSK3- binding peptide from FRAT1 selectively inhibits the GSK3-catalysed phosphorylation of axin and beta-catenin. **FEBS Lett**. 458, 247-51.

65. Peifer M, **Polakis P**. 2000 Wnt signaling in oncogenesis and embryogenesis--a look outside the nucleus. **Science** 287,1606-9.

66. **Polakis P.** 2000 Wnt signaling and cancer. **Genes Dev**;14, 1837-1851.

67. Spink KE, **Polakis P**, Weis WI 2000 Structural basis of the Axin-adenomatous polyposis coli interaction. **EMBO J** 19, 2270-2279.

68. Szeto , W., Jiang, W., Tice, D.A., Rubinfeld, B., Hollingshead, P.G., Fong, S.E., Dugger, D.L., Pham, T., Yansura, D.E., Wong, T.A., Grimaldi, J.C., Corpuz, R.T., Singh J.S., Frantz, G.D., Devaux, B., Crowley, C.W., Schwall, R.H., Eberhard, D.A., Rastelli, L., **Polakis, P.** and Pennica, D. 2001 Overexpression of the Retinoic Acid-Responsive Gene Stra6 in Human Cancers and its Synergistic Induction by Wnt-1 and Retinoic Acid. **Cancer Res** 61, 4197-4204.

69. Rubinfeld B, Tice DA, **Polakis P**. 2001 Axin dependent phosphorylation of the adenomatous polyposis coli protein mediated by casein kinase 1 epsilon. **J Biol Chem** 276, 39037-39045.

70. **Polakis P**. 2001 More than one way to skin a catenin. **Cell** 2001 105, 563-566.

71. Tice DA, Soloviev I, **Polakis P**. 2002 Activation of the Wnt Pathway Interferes withSerum Response Element-driven Transcription of Immediate Early Genes. **J Biol. Chem**. 277, 6118-6123.

72. Tice DA, Szeto W, Soloviev I, Rubinfeld B, Fong SE, Dugger DL, Winer J,

Williams PM, Wieand D, Smith V, Schwall RH, Pennnica D, **Polakis P.** 2002 Synergistic activation of tumor antigens by wnt-1 signaling and retinoic acid revealed by gene expression profiling. **J Biol Chem**. 277,14329-14335.

73. **Polakis, P.** 2002 Casein kinase I: A wnt'er of disconnect. **Curr. Biol.** 12, R499.

74. Mao,W. , Luis, E., Ross, S., Silva, J., Tan, C., Crowley, C., Chui, C., Franz, G., Senter, P., Koeppen, H., **Polakis, P.** 2004 EphB2 as a therapeutic antibody drug target for the treatment of colorectal cancer. **Cancer Res.** 64, 781-788.

75. Shibamoto, S., Winer, J., Williams, M., Polakis, P. 2003 A Blockade in Wnt signaling is activated following the differentiation of F9 teratocarcinoma cells. **Exp. Cell Res.** 29211-20.

76. Zhang Y, Eberhard DA, Frantz GD, Dowd P, Wu TD, Zhou Y, Watanabe C, Luoh SM, **Polakis P,** Hillan KJ, Wood WI, Zhang Z. 2004 GEPIS--quantitative gene expression profiling in normal and cancer tissues. **Bioinformatics**, April 8

Genome-wide Study of Gene Copy Numbers, Transcripts, and Protein Levels in Pairs of Non-invasive and Invasive Human Transitional Cell Carcinomas*

Torben F. Ørntoft‡§, Thomas Thykjaer¶, Frederic M. Waldman‖, Hans Wolf**, and Julio E. Celis‡‡

Gain and loss of chromosomal material is characteristic of bladder cancer, as well as malignant transformation in general. The consequences of these changes at both the transcription and translation levels is at present unknown partly because of technical limitations. Here we have attempted to address this question in pairs of non-invasive and invasive human bladder tumors using a combination of technology that included comparative genomic hybridization, high density oligonucleotide array-based monitoring of transcript levels (5600 genes), and high resolution two-dimensional gel electrophoresis. The results showed that there is a gene dosage effect that in some cases superimposes on other regulatory mechanisms. This effect depended ($p < 0.015$) on the magnitude of the comparative genomic hybridization change. In general (18 of 23 cases), chromosomal areas with more than 2-fold gain of DNA showed a corresponding increase in mRNA transcripts. Areas with loss of DNA, on the other hand, showed either reduced or unaltered transcript levels. Because most proteins resolved by two-dimensional gels are unknown it was only possible to compare mRNA and protein alterations in relatively few cases of well focused abundant proteins. With few exceptions we found a good correlation ($p < 0.005$) between transcript alterations and protein levels. The implications, as well as limitations, of the approach are discussed. ***Molecular & Cellular Proteomics 1:37–45, 2002.***

Aneuploidy is a common feature of most human cancers (1), but little is known about the genome-wide effect of this phenomenon at both the transcription and translation levels. High throughput array studies of the breast cancer cell line BT474 has suggested that there is a correlation between DNA copy numbers and gene expression in highly amplified areas (2), and studies of individual genes in solid tumors have revealed a good correlation between gene dose and mRNA or protein levels in the case of c-erb-B2, *cyclin d1*, *ems1*, and N-myc (3–5). However, a high cyclin D1 protein expression has been observed without simultaneous amplification (4), and a low level of c-myc copy number increase was observed without concomitant c-myc protein overexpression (6).

In human bladder tumors, karyotyping, fluorescent *in situ* hybridization, and comparative genomic hybridization (CGH)[1] have revealed chromosomal aberrations that seem to be characteristic of certain stages of disease progression. In the case of non-invasive pTa transitional cell carcinomas (TCCs), this includes loss of chromosome 9 or parts of it, as well as loss of Y in males. In minimally invasive pT1 TCCs, the following alterations have been reported: 2q−, 11p−, 1q+, 11q13+, 17q+, and 20q+ (7–12). It has been suggested that these regions harbor tumor suppressor genes and oncogenes; however, the large chromosomal areas involved often contain many genes, making meaningful predictions of the functional consequences of losses and gains very difficult.

In this investigation we have combined genome-wide technology for detecting genomic gains and losses (CGH) with gene expression profiling techniques (microarrays and proteomics) to determine the effect of gene copy number on transcript and protein levels in pairs of non-invasive and invasive human bladder TCCs.

EXPERIMENTAL PROCEDURES

Material—Bladder tumor biopsies were sampled after informed consent was obtained and after removal of tissue for routine pathology examination. By light microscopy tumors 335 and 532 were staged by an experienced pathologist as pTa (superficial papillary),

From the ‡Department of Clinical Biochemistry, Molecular Diagnostic Laboratory and **Department of Urology, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark, ¶AROS Applied Biotechnology ApS, Gustav Wiedsvej 10, DK-8000 Aarhus C, Denmark, ‖UCSF Cancer Center and Department of Laboratory Medicine, University of California, San Francisco, CA 94143-0808, and ‡‡Institute of Medical Biochemistry and Danish Centre for Human Genome Research, Ole Worms Allé 170, Aarhus University, DK-8000 Aarhus C, Denmark

Received, September 26, 2001, and in revised form, November 7, 2001

Published, MCP Papers in Press, November 13, 2001, DOI 10.1074/mcp.M100019-MCP200

[1] The abbreviations used are: CGH, comparative genomic hybridization; TCC, transitional cell carcinoma; LOH, loss of heterozygosity; PA-FABP, psoriasis-associated fatty acid-binding protein; 2D, two-dimensional.

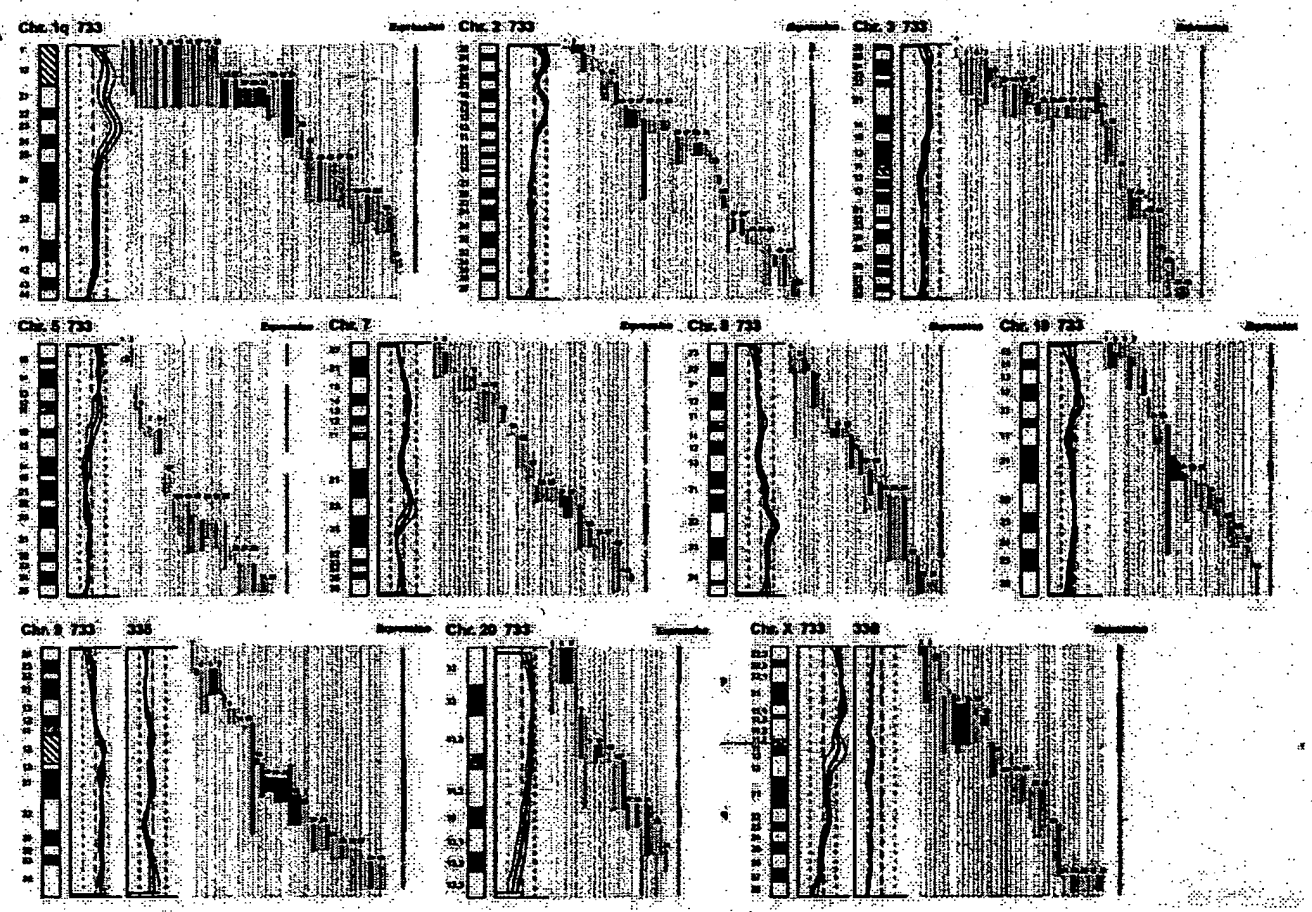


FIG. 1. **DNA copy number and mRNA expression level.** Shown from *left* to *right* are chromosome (*Chr.*), CGH profiles, gene location and expression level of specific genes, and overall expression level along the chromosome. *A*, expression of mRNA in invasive tumor 733 as compared with the non-invasive counterpart tumor 335. *B*, expression of mRNA in invasive tumor 827 compared with the non-invasive counterpart tumor 532. The average fluorescent signal ratio between tumor DNA and normal DNA is shown along the length of the chromosome (*left*). The *bold curve* in the ratio profile represents a mean of four chromosomes and is surrounded by *thin curves* indicating one standard deviation. The *central vertical line* (*broken*) indicates a ratio value of 1 (no change), and the *vertical lines* next to it (*dotted*) indicate a ratio of 0.5 (*left*) and 2.0 (*right*). In chromosomes where the non-invasive tumor 335 used for comparison showed alterations in DNA content, the ratio profile of that chromosome is shown to the *right* of the invasive tumor profile. The *colored bars* represents one gene each, identified by the running *numbers above* the *bars* (the name of the gene can be seen at www.MDL.DK/sdata.html). The *bars* indicate the purported location of the gene, and the *colors* indicate the expression level of the gene in the invasive tumor compared with the non-invasive counterpart; >2-fold increase (*black*), >2-fold decrease (*blue*), no significant change (*orange*). The *bar* to the *far right*, entitled *Expression* shows the resulting change in expression along the chromosome; the *colors* indicate that at least half of the genes were up-regulated (*black*), at least half of the genes down-regulated (*blue*), or more than half of the genes are unchanged (*orange*). If a gene was absent in one of the samples and present in another, it was regarded as more than a 2-fold change. A 2-fold level was chosen as this corresponded to one standard deviation in a double determination of ~1800 genes. Centromeres and heterochromatic regions were excluded from data analysis.

grade I and II, respectively, tumors 733 and 827 were staged as pT1 (invasive into submucosa), 733 was staged as solid, and 827 was staged as papillary, both grade III.

mRNA Preparation—Tissue biopsies, obtained fresh from surgery, were embedded immediately in a sodium-guanidinium thiocyanate solution and stored at −80 °C. Total RNA was isolated using the RNAzol B RNA isolation method (WAK-Chemie Medical GMBH). poly(A)$^+$ RNA was isolated by an oligo(dT) selection step (Oligotex mRNA kit; Qiagen).

cRNA Preparation—1 μg of mRNA was used as starting material. The first and second strand cDNA synthesis was performed using the SuperScript® choice system (Invitrogen) according to the manufacturer's instructions but using an oligo(dT) primer containing a T7 RNA polymerase binding site. Labeled cRNA was prepared using the MEGAscrip® *in vitro* transcription kit (Ambion). Biotin-labeled CTP and UTP (Enzo) was used, together with unlabeled NTPs in the reaction. Following the *in vitro* transcription reaction, the unincorporated nucleotides were removed using RNeasy columns (Qiagen).

Array Hybridization and Scanning—Array hybridization and scanning was modified from a previous method (13). 10 μg of cRNA was fragmented at 94 °C for 35 min in buffer containing 40 mM Tris acetate, pH 8.1, 100 mM KOAc, 30 mM MgOAc. Prior to hybridization, the fragmented cRNA in a 6× SSPE-T hybridization buffer (1 M NaCl, 10 mM Tris, pH 7.6, 0.005% Triton), was heated to 95 °C for 5 min, subsequently cooled to 40 °C, and loaded onto the Affymetrix probe array cartridge. The probe array was then incubated for 16 h at 40 °C at constant rotation (60 rpm). The probe array was exposed to 10 washes in 6× SSPE-T at 25 °C followed by 4 washes in 0.5× SSPE-T at 50 °C. The biotinylated cRNA was stained with a streptavidin-phycoerythrin conjugate, 10 μg/ml (Molecular Probes) in 6× SSPE-T

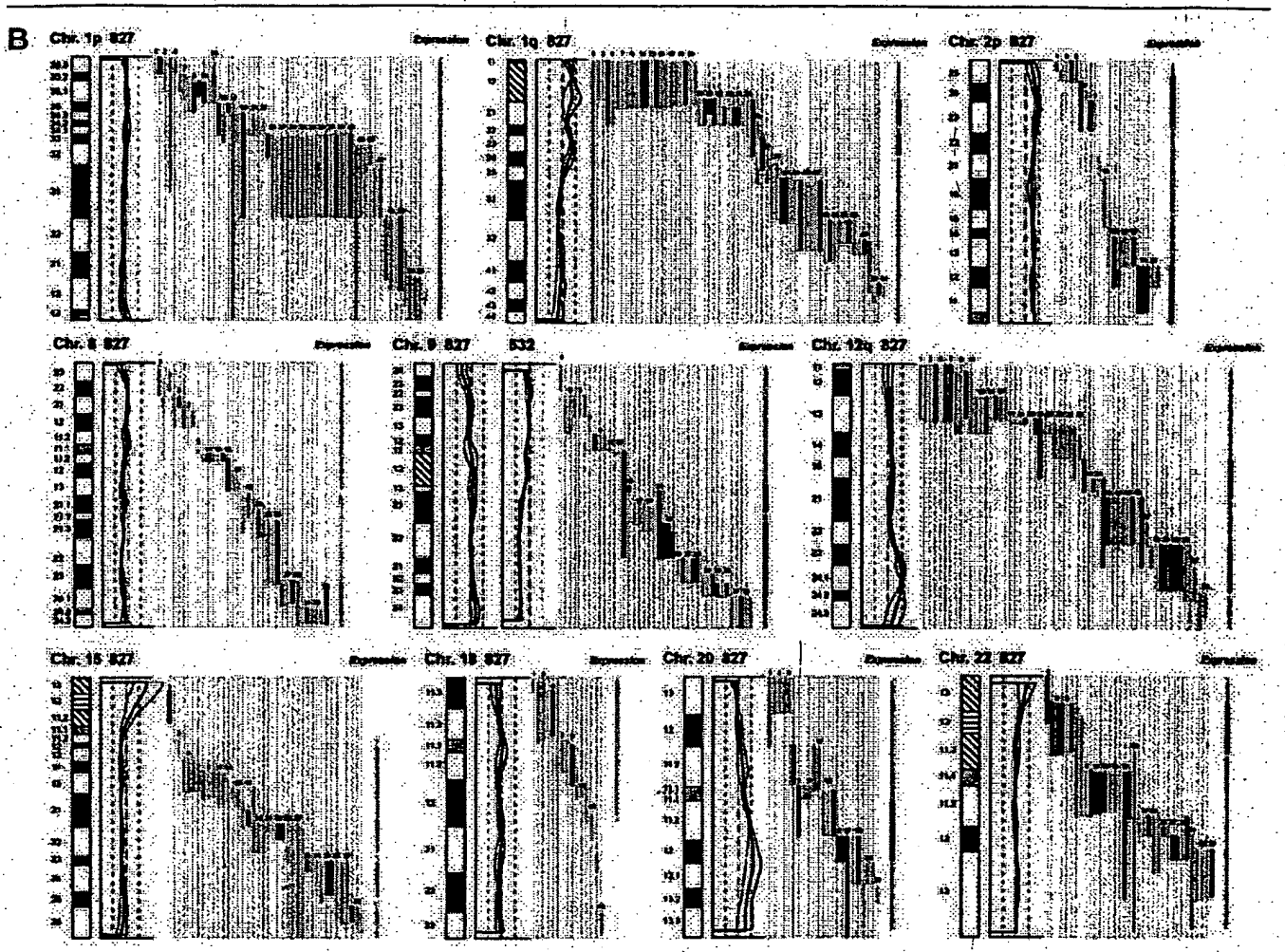


FIG. 1—*continued*

for 30 min at 25 °C followed by 10 washes in 6× SSPE-T at 25 °C. The probe arrays were scanned at 560 nm using a confocal laser scanning microscope (made for Affymetrix by Hewlett-Packard). The readings from the quantitative scanning were analyzed by Affymetrix gene expression analysis software.

Microsatellite Analysis—Microsatellite Analysis was performed as described previously (14). Microsatellites were selected by use of www.ncbi.nlm.nih.gov/genemap98, and primer sequences were obtained from the genome data base at www.gdb.org. DNA was extracted from tumor and blood and amplified by PCR in a volume of 20 μl for 35 cycles. The amplicons were denatured and electrophoresed for 3 h in an ABI Prism 377. Data were collected in the Gene Scan program for fragment analysis. Loss of heterozygosity was defined as less than 33% of one allele detected in tumor amplicons compared with blood.

Proteomic Analysis—TCCs were minced into small pieces and homogenized in a small glass homogenizer in 0.5 ml of lysis solution. Samples were stored at −20 °C until use. The procedure for 2D gel electrophoresis has been described in detail elsewhere (15, 16). Gels were stained with silver nitrate and/or Coomassie Brilliant Blue. Proteins were identified by a combination of procedures that included microsequencing, mass spectrometry, two-dimensional gel Western immunoblotting, and comparison with the master two-dimensional gel image of human keratinocyte proteins; see biobase.dk/cgi-bin/celis.

CGH—Hybridization of differentially labeled tumor and normal DNA to normal metaphase chromosomes was performed as described previously (10). Fluorescein-labeled tumor DNA (200 ng), Texas Red-labeled reference DNA (200 ng), and human Cot-1 DNA (20 μg) were denatured at 37 °C for 5 min and applied to denatured normal metaphase slides. Hybridization was at 37 °C for 2 days. After washing, the slides were counterstained with 0.15 μg/ml 4,6-diamidino-2-phenylindole in an anti-fade solution. A second hybridization was performed for all tumor samples using fluorescein-labeled reference DNA and Texas Red-labeled tumor DNA (inverse labeling) to confirm the aberrations detected during the initial hybridization. Each CGH experiment also included a normal control hybridization using fluorescein- and Texas Red-labeled normal DNA. Digital image analysis was used to identify chromosomal regions with abnormal fluorescence ratios, indicating regions of DNA gains and losses. The average green:red fluorescence intensity ratio profiles were calculated using four images of each chromosome (eight chromosomes total) with normalization of the green:red fluorescence intensity ratio for the entire metaphase and background correction. Chromosome identification was performed based on 4,6-diamidino-2-phenylindole banding patterns. Only images showing uniform high intensity fluorescence with minimal background staining were analyzed. All centromeres, p arms of acrocentric chromosomes, and heterochromatic regions were excluded from the analysis.

RESULTS

Comparative Genomic Hybridization—The CGH analysis identified a number of chromosomal gains and losses in the

TABLE I

Correlation between alterations detected by CGH and by expression monitoring

Top, CGH used as independent variable (if CGH alteration – what expression ratio was found); bottom, altered expression used as independent variable (if expression alteration – what CGH deviation was found).

CGH alterations	Tumor 733 vs. 335 Expression change clusters	Concordance	CGH alterations	Tumor 827 vs. 532 Expression change clusters	Concordance
13 Gain	10 Up-regulation 0 Down-regulation 3 No change	77%	10 Gain	8 Up-regulation 0 Down-regulation 2 No change	80%
10 Loss	1 Up-regulation 5 Down-regulation 4 No change	50%	12 Loss	3 Up-regulation 2 Down regulation 7 No change	17%

Expression change clusters	Tumor 733 vs. 335 CGH alterations	Concordance	Expression change clusters	Tumor 827 vs. 532 CGH alterations	Concordance
16 Up-regulation	11 Gain 2 Loss 3 No change	69%	17 Up-regulation	10 Gain 5 Loss 2 No change	59%
21 Down-regulation	1 Gain 8 Loss 12 No change	38%	9 Down-regulation	0 Gain 3 Loss 6 No change	33%
15 No change	3 Gain 3 Loss 9 No change	60%	21 No change	1 Gain 3 Loss 17 No change	81%

two invasive tumors (stage pT1, TCCs 733 and 827), whereas the two non-invasive papillomas (stage pTa, TCCs 335 and 532) showed only 9p–, 9q22-q33–, and X–, and 7+, 9q–, and Y–, respectively. Both invasive tumors showed changes (1q22-24+, 2q14.1-qter–, 3q12-q13.3–, 6q12-q22–, 9q34+, 11q12-q13+, 17+, and 20q11.2-q12+) that are typical for their disease stage, as well as additional alterations, some of which are shown in Fig. 1. Areas with gains and losses deviated from the normal copy number to some extent, and the average numerical deviation from normal was 0.4-fold in the case of TCC 733 and 0.3-fold for TCC 827. The largest changes, amounting to at least a doubling of chromosomal content, were observed at 1q23 in TCC 733 (Fig. 1*A*) and 20q12 in TCC 827 (Fig. 1*B*).

mRNA Expression in Relation to DNA Copy Number—The mRNA levels from the two invasive tumors (TCCs 827 and 733) were compared with the two non-invasive counterparts (TCCs 532 and 335). This was done in two separate experiments in which we compared TCCs 733 to 335 and 827 to 532, respectively, using two different scaling settings for the arrays to rule out scaling as a confounding parameter. Approximately 1,800 genes that yielded a signal on the arrays were searched in the Unigene and Genemap data bases for chromosomal location, and those with a known location (1096) were plotted as bars covering their purported locus. In that way it was possible to construct a graphic presentation of DNA copy number and relative mRNA levels along the individual chromosomes (Fig. 1).

For each mRNA a ratio was calculated between the level in the invasive *versus* the non-invasive counterpart. Bars, which represent chromosomal location of a gene, were color-coded according to the expression ratio, and only differences larger than 2-fold were regarded as informative (Fig. 1). The density of genes along the chromosomes varied, and areas containing only one gene were excluded from the calculations. The resolution of the CGH method is very low, and some of the outlier data may be because of the fact that the boundaries of the chromosomal aberrations are not known at high resolution.

Two sets of calculations were made from the data. For the first set we used CGH alterations as the independent variable and estimated the frequency of expression alterations in these chromosomal areas. In general, areas with a strong gain of chromosomal material contained a cluster of genes having increased mRNA expression. For example, both chromosomes 1q21-q25, 2p and 9q, showed a relative gain of more than 100% in DNA copy number that was accompanied by increased mRNA expression levels in the two tumor pairs (Fig. 1). In most cases, chromosomal gains detected by CGH were accompanied by an increased level of transcripts in both TCCs 733 (77%) and 827 (80%) (Table I, *top*). Chromosomal losses, on the other hand, were not accompanied by decreased expression in several cases, and were often registered as having unaltered RNA levels (Table I, *top*). The inability to detect RNA expression changes in these cases was not because of fewer genes mapping to the lost regions (data not shown).

In the second set of calculations we selected expression alterations above 2-fold as the independent variable and estimated the frequency of CGH alterations in these areas. As above, we found that increased transcript expression correlated with gain of chromosomal material (TCC 733, 69% and TCC 827, 59%), whereas reduced expression was often detected in areas with unaltered CGH ratios (Table I, *bottom*). Furthermore, as a control we looked at areas with no alter-

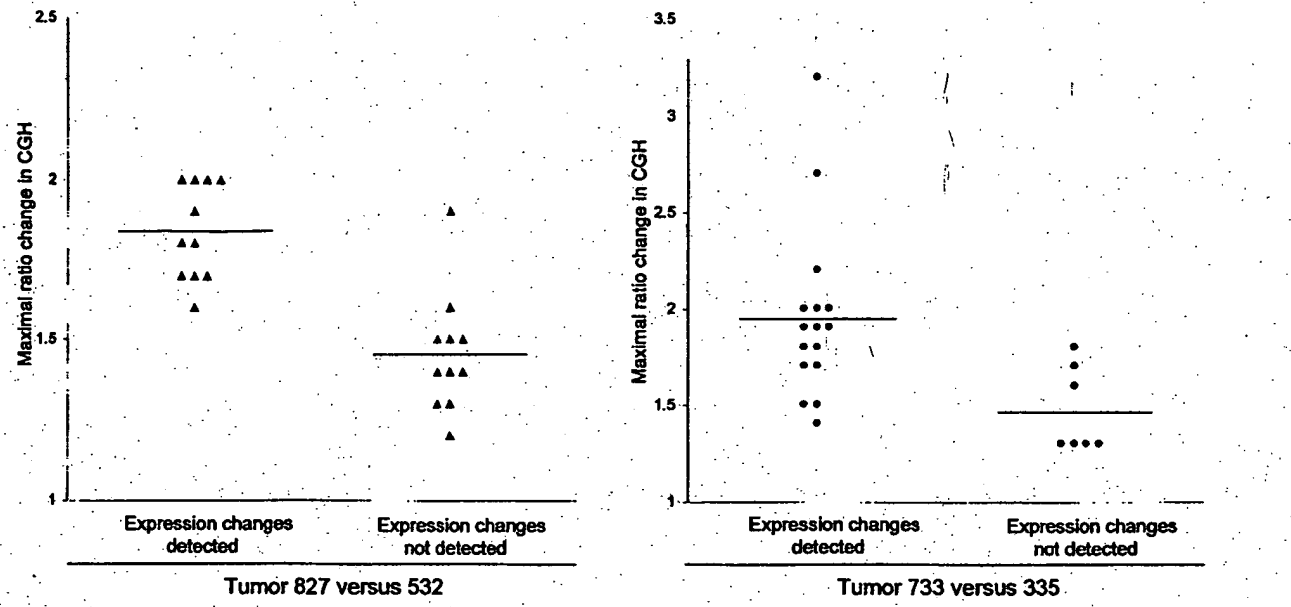


FIG. 2. **Correlation between maximum CGH aberration and the ability to detect expression change by oligonucleotide array monitoring.** The aberration is shown as a numerical -fold change in ratio between invasive tumors 827 (▲) and 733 (♦) and their non-invasive counterparts 532 and 335. The expression change was taken from the *Expression* line to the *right* in Fig. 1, which depicts the resulting expression change for a given chromosomal region. At least half of the mRNAs from a given region have to be either up- or down-regulated to be scored as an expression change. All chromosomal arms in which the CGH ratio plus or minus one standard deviation was outside the ratio value of one were included.

ation in expression. No alteration was detected by CGH in most of these areas (TCC 733, 60% and TCC 827, 81%; see Table I, *bottom*). Because the ability to observe reduced or increased mRNA expression clustering to a certain chromosomal area clearly reflected the extent of copy number changes, we plotted the maximum CGH aberrations in the regions showing CGH changes against the ability to detect a change in mRNA expression as monitored by the oligonucleotide arrays (Fig. 2). For both tumors TCC 733 ($p < 0.015$) and TCC 827 ($p < 0.00003$) a highly significant correlation was observed between the level of CGH ratio change (reflecting the DNA copy number) and alterations detected by the array based technology (Fig. 2). Similar data were obtained when areas with altered expression were used as independent variables. These areas correlated best with CGH when the CGH ratio deviated 1.6- to 2.0-fold (Table I, *bottom*) but mostly did not at lower CGH deviations. These data probably reflect that loss of an allele may only lead to a 50% reduction in expression level, which is at the cut-off point for detection of expression alterations. Gain of chromosomal material can occur to a much larger extent.

Microsatellite-based Detection of Minor Areas of Losses—In TCC 733, several chromosomal areas exhibiting DNA amplification were preceded or followed by areas with a normal CGH but reduced mRNA expression (see Fig. 1, TCC 733 chromosome 1q32, 2p21, and 7q21 and q32, 9q34, and 10q22). To determine whether these results were because of undetected loss of chromosomal material in these regions or because of other non-structural mechanisms regulating transcription, we examined two microsatellites positioned at chromosome 1q25–32 and two at chromosome 2p22. Loss of heterozygosity (LOH) was found at both 1q25 and at 2p22 indicating that minor deleted areas were not detected with the resolution of CGH (Fig. 3). Additionally, chromosome 2p in TCC 733 showed a CGH pattern of gain/no change/gain of DNA that correlated with transcript increase/decrease/increase. Thus, for the areas showing increased expression there was a correlation with the DNA copy number alterations (Fig. 1*A*). As indicated above, the mRNA decrease observed in the middle of the chromosomal gain was because of LOH, implying that one of the mechanisms for mRNA down-regulation may be regions that have undergone smaller losses of chromosomal material. However, this cannot be detected with the resolution of the CGH method.

In both TCC 733 and TCC 827, the telomeric end of chromosome 11p showed a normal ratio in the CGH analysis; however, clusters of five and three genes, respectively, lost their expression. Two microsatellites (D11S1760, D11S922) positioned close to MUC2, IGF2, and cathepsin D indicated LOH as the most likely mechanism behind the loss of expression (data not shown).

A reduced expression of mRNA observed in TCC 733 at chromosomes 3q24, 11p11, 12p12.2, 12q21.1, and 16q24 and in TCC 827 at chromosome 11p15.5, 12p11, 15q11.2, and 18q12 was also examined for chromosomal losses using microsatellites positioned as close as possible to the gene loci

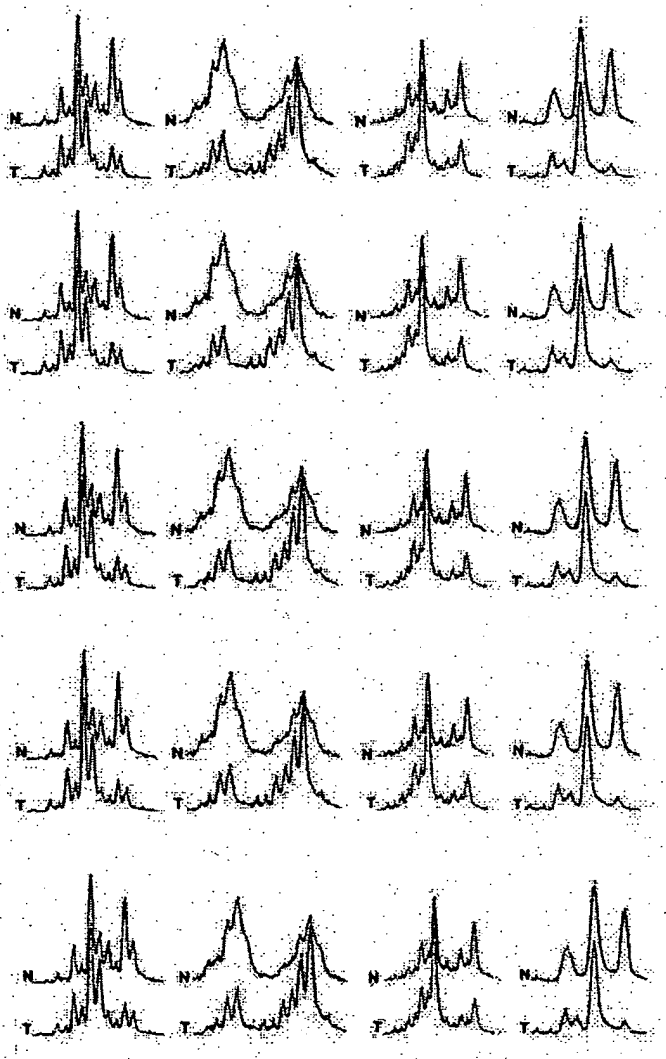


FIG. 3. **Microsatellite analysis of loss of heterozygosity.** Tumor 733 showing loss of heterozygosity at chromosome 1q25, detected (*a*) by D1S215 close to Hu class I histocompatibility antigen (gene number 38 in Fig. 1), (*b*) by D1S2735 close to cathepsin E (gene number 41 in Fig. 1), and (*c*) at chromosome 2p23 by D2S2251 close to general β-spectrin (gene number 11 on Fig. 1) and of (*d*) tumor 827 showing loss of heterozygosity at chromosome 18q12 by S18S1118 close to mitochondrial 3-oxoacyl-coenzyme A thiolase (gene number 12 in Fig. 1). The *upper curves* show the electropherogram obtained from normal DNA from leukocytes (*N*), and the *lower curves* show the electropherogram from tumor DNA (*T*). In all cases one allele is partially lost in the tumor amplicon.

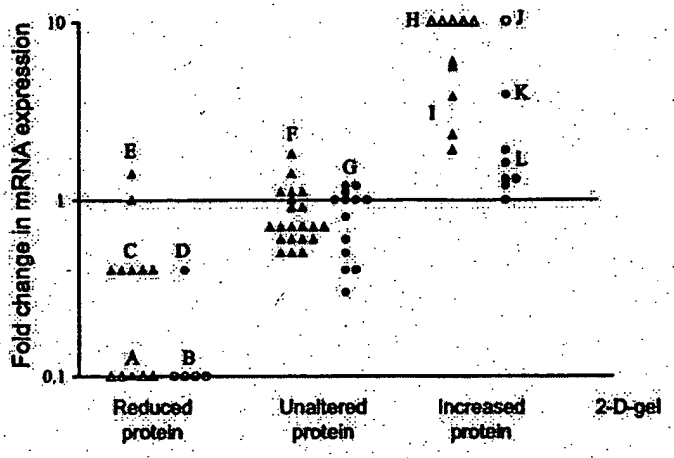


FIG. 4. **Correlation between protein levels as judged by 2D-PAGE and transcript ratio.** For comparison proteins were divided in three groups, unaltered in level or up- or down-regulated (*horizontal axis*). The mRNA ratio as determined by oligonucleotide arrays was plotted for each gene (*vertical axis*). ▲, mRNAs that were scored as present in both tumors used for the ratio calculation; △, mRNAs that were scored as absent in the invasive tumors (along *horizontal axis*) or as absent in non-invasive reference (*top* of figure). Two different scalings were used to exclude scaling as a confounder, TCCs 827 and 532 (▲△) were scaled with background suppression, and TCCs 733 and 335 (●○) were scaled without suppression. Both comparisons showed highly significant ($p < 0.005$) differences in mRNA ratios between the groups. Proteins shown were as follows: Group *A* (from *left*), phosphoglucomutase 1, glutathione transferase class μ number 4, fatty acid-binding protein homologue, cytokeratin 15, and cytokeratin 13; *B* (from *left*), fatty acid-binding protein homologue, 28-kDa heat shock protein, cytokeratin 13, and calcyclin; *C* (from *left*), α-enolase, hnRNP B1, 28-kDa heat shock protein, 14–3-3-ε, and pre-mRNA splicing factor; *D*, mesothelial keratin K7 (type II); *E* (from *top*), glutathione *S*-transferase-π and mesothelial keratin K7 (type II); *F* (from *top* and *left*), adenylyl cyclase-associated protein, E-cadherin, keratin 19, calgizzarin, phosphoglycerate mutase, annexin IV, cytoskeletal γ-actin, hnRNP A1, integral membrane protein calnexin (IP90), hnRNP H, brain-type clathrin light chain-a, hnRNP F, 70-kDa heat shock protein, heterogeneous nuclear ribonucleoprotein A/B, translationally controlled tumor protein, liver glyceraldehyde-3-phosphate dehydrogenase, keratin 8, aldehyde reductase, and Na,K-ATPase β-1 subunit; *G*, (from *top* and *left*), TCP20, calgizzarin, 70-kDa heat shock protein, calnexin, hnRNP H, cytokeratin 15, ATP synthase, keratin 19, triosephosphate isomerase, hnRNP F, liver glyceraldehyde-3-phosphatase dehydrogenase, glutathione *S*-transferase-π, and keratin 8; *H* (from *left*), plasma gelsolin, autoantigen calreticulin, thioredoxin, and NAD+-dependent 15 hydroxyprostaglandin dehydrogenase; *I* (from *top*), prolyl 4-hydroxylase β-subunit, cytokeratin 20, cytokeratin 17, prohibition, and fructose 1,6-biphosphatase; *J* annexin II; *K*, annexin IV; *L* (from *top* and *left*), 90-kDa heat shock protein, prolyl 4-hydroxylase β-subunit, α-enolase, GRP 78, cyclophilin, and cofilin.

showing reduced mRNA transcripts. Only the microsatellite positioned at 18q12 showed LOH (Fig. 3), suggesting that transcriptional down-regulation of genes in the other regions may be controlled by other mechanisms.

Relation between Changes in mRNA and Protein Levels—2D-PAGE analysis, in combination with Coomassie Brilliant Blue and/or silver staining, was carried out on all four tumors using fresh biopsy material. 40 well resolved abundant known proteins migrating in areas away from the edges of the pH gradient, and having a known chromosomal location, were selected for analysis in the TCC pair 827/532. Proteins were identified by a combination of methods (see "Experimental Procedures"). In general there was a highly significant correlation ($p < 0.005$) between mRNA and protein alterations (Fig. 4). Only one gene showed disagreement between transcript alteration and protein alteration. Except for a group of cyto-

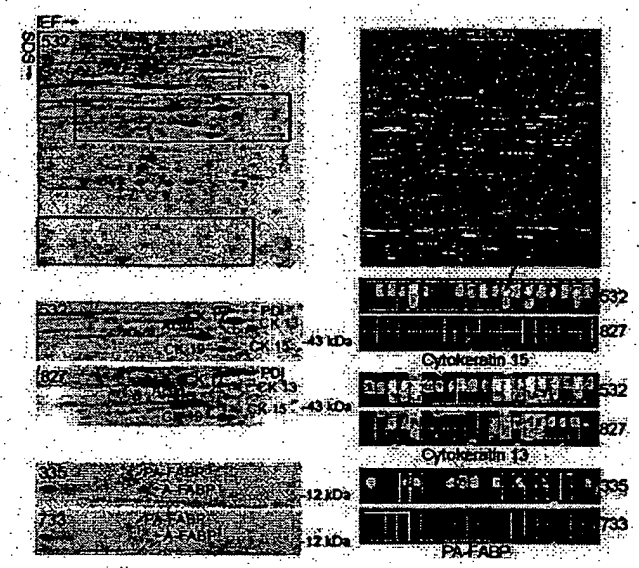


FIG. 5. **Comparison of protein and transcript levels in invasive and non-invasive TCCs.** The *upper part* of the figure shows a 2D gel (*left*) and the oligonucleotide array (*right*) of TCC 532. The *red rectangles* on the *upper gel* highlight the areas that are compared below. Identical areas of 2D gels of TCCs 532 and 827 are shown below. Clearly, cytokeratins 13 and 15 are strongly down-regulated in TCC 827 (*red annotation*). The tile on the array containing probes for cytokeratin 15 is enlarged *below* the array (*red arrow*) from TCC 532 and is compared with TCC 827. The *upper row* of *squares* in each tile corresponds to perfect match probes; the *lower row* corresponds to mismatch probes containing a mutation (used for correction for unspecific binding). Absence of signal is depicted as *black*, and the higher the signal the lighter the color. A high transcript level was detected in TCC 532 (6151 units) whereas a much lower level was detected in TCC 827 (absence of signals). For cytokeratin 13, a high transcript level was also present in TCC 532 (15659 units), and a much lower level was present in TCC 827 (623 units). The 2D gels at the *bottom* of the figure (*left*) show levels of PA-FABP and adipocyte-FABP in TCCs 335 and 733 (invasive), respectively. Both proteins are down-regulated in the invasive tumor. To the *right* we show the array tiles for the PA-FABP transcript. A medium transcript level was detected in the case of TCC 335 (1277 units) whereas very low levels were detected in TCC 733 (166 units). *IEF*, isoelectric focusing.

keratins encoded by genes on chromosome 17 (Fig. 5) the analyzed proteins did not belong to a particular family. 26 well focused proteins whose genes had a know chromosomal location were detected in TCCs 733 and 335, and of these 19 correlated ($p < 0.005$) with the mRNA changes detected using the arrays (Fig. 4). For example, PA-FABP was highly expressed in the non-invasive TCC 335 but lost in the invasive counterpart (TCC 733; see Fig. 5). The smaller number of proteins detected in both 733 and 335 was because of the smaller size of the biopsies that were available.

11 chromosomal regions where CGH showed aberrations that corresponded to the changes in transcript levels also showed corresponding changes in the protein level (Table II). These regions included genes that encode proteins that are found to be frequently altered in bladder cancer, namely cytokeratins 17 and 20, annexins II and IV, and the fatty acid-binding proteins PA-FABP and FBP1. Four of these proteins were encoded by genes in chromosome 17q, a frequently amplified chromosomal area in invasive bladder cancers.

DISCUSSION

Most human cancers have abnormal DNA content, having lost some chromosomal parts and gained others. The present study provides some evidence as to the effect of these gains and losses on gene expression in two pairs of non-invasive and invasive TCCs using high throughput expression arrays and proteomics, in combination with CGH. In general, the results showed that there is a clear individual regulation of the mRNA expression of single genes, which in some cases was superimposed by a DNA copy number effect. In most cases, genes located in chromosomal areas with gains often exhibited increased mRNA expression, whereas areas showing losses showed either no change or a reduced mRNA expression. The latter might be because of the fact that losses most often are restricted to loss of one allele, and the cut-off point for detection of expression alterations was a 2-fold change, thus being at the border of detection. In several cases, how-

TABLE II
Proteins whose expression level correlates with both mRNA and gene dose changes

Protein	Chromosomal location	Tumor TCC	CGH alteration	Transcript alteration[a]	Protein alteration
Annexin II	1q21	733	Gain	Abs to Pres[a]	Increase
Annexin IV	2p13	733	Gain	3.9-Fold up	Increase
Cytokeratin 17	17q12-q21	827	Gain	3.8-Fold up	Increase
Cytokeratin 20	17q21.1	827	Gain	5.6-Fold up	Increase
(PA-)FABP	8q21.2	827	Loss	10-Fold down	Decrease
FBP1	9q22	827	Gain	2.3-Fold up	Increase
Plasma gelsolin	9q31	827	Gain	Abs to Pres	Increase
Heat shock protein 28	15q12-q13	827	Loss	2.5-Fold up	Decrease
Prohibitin	17q21	827/733	Gain	3.7-/2.5-Fold up[b]	Increase
Prolyl-4-hydroxyl	17q25	827/733	Gain	5.7-/1.6-Fold up	Increase
hnRNPB1	7p15	827	Loss	2.5-Fold down	Decrease

[a] Abs, absent; Pres, present.
[b] In cases where the corresponding alterations were found in both TCCs 827 and 733 these are shown as 827/733.

ever, an increase or decrease in DNA copy number was associated with *de novo* occurrence or complete loss of transcript, respectively. Some of these transcripts could not be detected in the non-invasive tumor but were present at relatively high levels in areas with DNA amplifications in the invasive tumors (*e.g.* in TCC 733 transcript from cellular ligand of annexin II gene (chromosome 1q21) from absent to 2670 arbitrary units; in TCC 827 transcript from small proline-rich protein 1 gene (chromosome 1q12-q21.1) from absent to 1326 arbitrary units). It may be anticipated from these data that significant clustering of genes with an increased expression to a certain chromosomal area indicates an increased likelihood of gain of chromosomal material in this area.

Considering the many possible regulatory mechanisms acting at the level of transcription, it seems striking that the gene dose effects were so clearly detectable in gained areas. One hypothetical explanation may lie in the loss of controlled methylation in tumor cells (17–19). Thus, it may be possible that in chromosomes with increased DNA copy numbers two or more alleles could be demethylated simultaneously leading to a higher transcription level, whereas in chromosomes with losses the remaining allele could be partly methylated, turning off the process (20, 21). A recent report has documented a ploidy regulation of gene expression in yeast, but in this case all the genes were present in the same ratio (22), a situation that is not analogous to that of cancer cells, which show marked chromosomal aberrations, as well as gene dosage effects.

Several CGH studies of bladder cancer have shown that some chromosomal aberrations are common at certain stages of disease progression, often occurring in more than 1 of 3 tumors. In pTa tumors, these include 9p−, 9q−, 1q+, Y− (2, 6), and in pT1 tumors, 2q−,11p−, 11q−, 1q+, 5p+, 8q+, 17q+, and 20q+ (2–4, 6, 7). The pTa tumors studied here showed similar aberrations such as 9p− and 9q22-q33− and 9q− and Y−, respectively. Likewise, the two minimal invasive pT1 tumors showed aberrations that are commonly seen at that stage, and TCC 827 had a remarkable resemblance to the commonly seen pattern of losses and gains, such as 1q22–24 amplification (seen in both tumors), 11q14-q22 loss, the latter often linked to 17 q+ (both tumors), and 1q+ and 9p−, often linked to 20q+ and 11 q13+ (both tumors) (7–9). These observations indicate that the pairs of tumors used in this study exhibit chromosomal changes observed in many tumors, and therefore the findings could be of general importance for bladder cancer.

Considering that the mapping resolution of CGH is of about 20 megabases it is only possible to get a crude picture of chromosomal instability using this technique. Occasionally, we observed reduced transcript levels close to or inside regions with increased copy numbers. Analysis of these regions by positioning heterozygous microsatellites as close as possible to the locus showing reduced gene expression revealed loss of heterozygosity in several cases. It seems likely that multiple and different events occur along each chromosomal arm and that the use of cDNA microarrays for analysis of DNA copy number changes will reach a resolution that can resolve these changes, as has recently been proposed (2). The outlier data were not more frequent at the boundaries of the CGH aberrations. At present we do not know the mechanism behind chromosomal aneuploidy and cannot predict whether chromosomal gains will be transcribed to a larger extent than the two native alleles. A mechanism as genetic imprinting has an impact on the expression level in normal cells and is often reduced in tumors. However, the relation between imprinting and gain of chromosomal material is not known.

We regard it as a strength of this investigation that we were able to compare invasive tumors to benign tumors rather than to normal urothelium, as the tumors studied were biologically very close and probably may represent successive steps in the progression of bladder cancer. Despite the limited amount of fresh tissue available it was possible to apply three different state of the art methods. The observed correlation between DNA copy number and mRNA expression is remarkable when one considers that different pieces of the tumor biopsies were used for the different sets of experiments. This indicate that bladder tumors are relatively homogenous, a notion recently supported by CGH and LOH data that showed a remarkable similarity even between tumors and distant metastasis (10, 23).

In the few cases analyzed, mRNA and protein levels showed a striking correspondence although in some cases we found discrepancies that may be attributed to translational regulation, post-translational processing, protein degradation, or a combination of these. Some transcripts belong to undertranslated mRNA pools, which are associated with few translationally inactive ribosomes; these pools, however, seem to be rare (24). Protein degradation, for example, may be very important in the case of polypeptides with a short half-life (*e.g.* signaling proteins). A poor correlation between mRNA and protein levels was found in liver cells as determined by arrays and 2D-PAGE (25), and a moderate correlation was recently reported by Ideker *et al.* (26) in yeast.

Interestingly, our study revealed a much better correlation between gained chromosomal areas and increased mRNA levels than between loss of chromosomal areas and reduced mRNA levels. In general, the level of CGH change determined the ability to detect a change in transcript. One possible explanation could be that by losing one allele the change in mRNA level is not so dramatic as compared with gain of material, which can be rather unlimited and may lead to a severalfold increase in gene copy number resulting in a much higher impact on transcript level. The latter would be much easier to detect on the expression arrays as the cut-off point was placed at a 2-fold level so as not to be biased by noise on the array. Construction of arrays with a better signal to noise ratio may in the future allow detection of lesser than 2-fold alterations in transcript levels, a feature that may facilitate the analysis of the effect of loss of chromosomal areas on transcript levels.

In eleven cases we found a significant correlation between DNA copy number, mRNA expression, and protein level. Four of these proteins were encoded by genes located at a frequently amplified area in chromosome 17q. Whether DNA copy number is one of the mechanisms behind alteration of these eleven proteins is at present unknown and will have to be proved by other methods using a larger number of samples. One factor making such studies complicated is the large extent of protein modification that occurs after translation, requiring immunoidentification and/or mass spectrometry to correctly identify the proteins in the gels.

In conclusion, the results presented in this study exemplify the large body of knowledge that may be possible to gather in the future by combining state of the art techniques that follow the pathway from DNA to protein (26). Here, we used a traditional chromosomal CGH method, but in the future high resolution CGH based on microarrays with many thousand radiation hybrid-mapped genes will increase the resolution and information derived from these types of experiments (2). Combined with expression arrays analyzing transcripts derived from genes with known locations, and 2D gel analysis to obtain information at the post-translational level, a clearer and more developed understanding of the tumor genome will be forthcoming.

Acknowledgments—We thank Mie Madsen, Hanne Steen, Inge Lis Thorsen, Hans Lund, Vikolaj Ørntoft, and Lynn Bjerke for technical help and Thomas Gingeras, Christine Harrington, and Morten Østergaard for valuable discussions.

* This work was supported by grants from The Danish Cancer Society, the University of Aarhus, Aarhus County, Novo Nordic, the Danish Biotechnology Program, the Frenkels Foundation, the John and Birthe Meyer Foundation, and NCI, National Institutes of Health Grant CA47537. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Dept. of Clinical Biochemistry, Molecular Diagnostic Laboratory, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark. Tel.: 45-89495100/45-86156201 (private); Fax: 45-89496018; E-mail: orntoft@kba.sks.au.dk.

REFERENCES

1. Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998) Genetic instabilities in human cancers. *Nature* **17**, 643–649
2. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* **23**, 41–46
3. de Cremoux, P., Martin, E. C., Vincent-Salomon, A., Dieras, V., Barbaroux, C., Liva, S., Pouillart, P., Sastre-Garau, X., and Magdelenat, H. (1999) Quantitative PCR analysis of c-erb B-2 (HER2/neu) gene amplification and comparison with p185(HER2/neu) protein expression in breast cancer drill biopsies. *Int. J. Cancer* **83**, 157–161
4. Brungier, P. P., Tamimi, Y., Shuuring, E., and Schalken, J. (1996) Expression of cyclin D1 and EMS1 in bladder tumors; relationship with chromosome 11q13 amplifications. *Oncogene* **12**, 1747–1753
5. Slavc, I., Ellenbogen, R., Jung, W. H., Vawter, G. F., Kretschmar, C., Grier, H., and Korf, B. R. (1990) *myc* gene amplification and expression in primary human neuroblastoma. *Cancer Res.* **50**, 1459–1463
6. Sauter, G., Carroll, P., Moch, H., Kallioniemi, A., Kerschmann, R., Narayan, P., Mihatsch, M. J., and Waldman, F. M. (1995) c-*myc* copy number gains in bladder cancer detected by fluorescence *in situ* hybridization. *Am J. Pathol.* **146**, 1131–1139
7. Richter, J., Jiang, F., Gorog, J. P., Sartorious, G., Egenter, C., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1997) Marked genetic differences between stage pTa and stage pT1 papillary bladder cancer detected by comparative genomic hybridization. *Cancer Res.* **57**, 2860–2864
8. Richter, J., Beffa, L., Wagner, U., Schraml, P., Gasser, T. C., Moch, H., Mihatsch, M. J., and Sauter, G. (1998) Patterns of chromosomal imbalances in advanced urinary bladder cancer detected by comparative genomic hybridization. *Am. J. Pathol.* **153**, 1615–1621
9. Bruch, J., Wohr, G., Hautmann, R., Mattfeldt, T., Bruderlein, S., Moller, P., Sauter, S., Hameister, H., Vogel, W., and Paiss, T. (1998) Chromosomal changes during progression of transitional cell carcinoma of the bladder and delineation of the amplified interval on chromosome arm 8q. *Genes Chromosomes Cancer* **23**, 167–174
10. Hovey, R. M., Chu, L., Balazs, M., De Vries, S., Moore, D., Sauter, G., Carroll, P. R., and Waldman, F. M. (1998) Genetic alterations in primary bladder cancers and their metastases. *Cancer Res.* **15**, 3555–3560
11. Simon, R., Burger, H., Brinkschmidt, C., Bocker, W., Hertle, L., and Terpe, H. J. (1998) Chromosomal aberrations associated with invasion in papillary superficial bladder cancer. *J. Pathol.* **185**, 345–351
12. Koo, S. H., Kwon, K. C., Ihm, C. H., Jeon, Y. M., Park, J. W., and Sul, C. K. (1999) Detection of genetic alterations in bladder tumors by comparative genomic hybridization and cytogenetic analysis. *Cancer Genet. Cytogenet.* **110**, 87–93
13. Wodicka, L., Dong, H., Mittmann, M., Ho, M. H., and Lockhart, D. J. (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **15**, 1359–1367
14. Christensen, M., Sunde, L., Bolund, L., and Orntoft, T. F. (1999) Comparison of three methods of microsatellite detection. *Scand. J. Clin. Lab. Invest.* **59**, 167–177
15. Celis, J. E., Ostergaard, M., Basse, B., Celis, A., Lauridsen, J. B., Ratz, G. P., Andersen, I., Hein, B., Wolf, H., Orntoft, T. F., and Rasmussen, H. H. (1996) Loss of adipocyte-type fatty acid binding protein and other protein biomarkers is associated with progression of human bladder transitional cell carcinomas. *Cancer Res.* **56**, 4782–4790
16. Celis, J. E., Ratz, G., Basse, B., Lauridsen, J. B., and Celis, A. (1994) in *Cell Biology: A Laboratory Handbook* (Celis, J. E., ed) Vol. 3, pp. 222–230, Academic Press, Orlando, FL
17. Ohlsson, R., Tycko, B., and Sapienza, C. (1998) Monoallelic expression: 'there can only be one'. *Trends Genet.* **14**, 435–438
18. Hollander, G. A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., Terhorst, C., Wishart, W., Golan, D. E., Bhan, A. K., and Burakoff, S. J. (1998) Monoallelic expression of the interleukin-2 locus. *Science* **279**, 2118–2121
19. Brannan, C. I., and Bartolomei, M. S. (1999) Mechanisms of genomic imprinting. *Curr. Opin. Genet. Dev.* **9**, 164–170
20. Ohlsson, R., Cui, H., He, L., Pfeifer, S., Malmikumpa, H., Jiang, S., Feinberg, A. P., and Hedborg, F. (1999) Mosaic allelic insulin-like growth factor 2 expression patterns reveal a link between Wilms' tumorigenesis and epigenetic heterogeneity. *Cancer Res.* **59**, 3889–3892
21. Cui, H., Hedborg, F., He, L., Nordenskjold, A., Sandstedt, B., Pfeifer-Ohlsson, S., and Ohlsson, R. (1997) Inactivation of H19, an imprinted and putative tumor repressor gene, is a preneoplastic event during Wilms' tumorigenesis. *Cancer Res.* **57**, 4469–4473
22. Galitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S., and Fink, G. R. (1999) Ploidy regulation of gene expression. *Science* **285**, 251–254
23. Tsao, J., Yatabe, Y., Marki, I. D., Hajyan, K., Jones, P. A., and Shibata, D. (2000) Bladder cancer genotype stability during clinical progression. *Genes Chromosomes Cancer* **29**, 26–32
24. Zong, Q., Schummer, M., Hood, L., and Morris, D. R. (1999) Messenger RNA translation state: the second dimension of high-throughput expression screening. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 10632–10636
25. Anderson, L., and Seilhamer, J. (1997) Comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**, 533–537
26. Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934

[CANCER RESEARCH 62, 6240–6245, November 1, 2002]

Impact of DNA Amplification on Gene Expression Patterns in Breast Cancer[1,2]

Elizabeth Hyman,[3] Päivikki Kauraniemi,[3] Sampsa Hautaniemi, Maija Wolf, Spyro Mousses, Ester Rozenblum, Markus Ringnér, Guido Sauter, Outi Monni, Abdel Elkahloun, Olli-P. Kallioniemi, and Anne Kallioniemi[4]

Howard Hughes Medical Institute-NIH Research Scholar, Bethesda, Maryland 20892 [E. H.]; Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892 [E. H., P. K., S. H., M. W., S. M., E. R., M. R., A. E., O. K., A. K.]; Laboratory of Cancer Genetics, Institute of Medical Technology, University of Tampere and Tampere University Hospital, FIN-33520 Tampere, Finland [P. K., A. K.]; Signal Processing Laboratory, Tampere University of Technology, FIN-33101 Tampere, Finland [S. H.]; Institute of Pathology, University of Basel, CH-4003 Basel, Switzerland [G. S.]; and Biomedicum Biochip Center, Helsinki University Hospital, Biomedicum Helsinki, FIN-00014 Helsinki, Finland [O. M.]

ABSTRACT

Genetic changes underlie tumor progression and may lead to cancer-specific expression of critical genes. Over 1100 publications have described the use of comparative genomic hybridization (CGH) to analyze the pattern of copy number alterations in cancer, but very few of the genes affected are known. Here, we performed high-resolution CGH analysis on cDNA microarrays in breast cancer and directly compared copy number and mRNA expression levels of 13,824 genes to quantitate the impact of genomic changes on gene expression. We identified and mapped the boundaries of 24 independent amplicons, ranging in size from 0.2 to 12 Mb. Throughout the genome, both high- and low-level copy number changes had a substantial impact on gene expression, with 44% of the highly amplified genes showing overexpression and 10.5% of the highly overexpressed genes being amplified. Statistical analysis with random permutation tests identified 270 genes whose expression levels across 14 samples were systematically attributable to gene amplification. These included most previously described amplified genes in breast cancer and many novel targets for genomic alterations, including the *HOXB7* gene, the presence of which in a novel amplicon at 17q21.3 was validated in 10.2% of primary breast cancers and associated with poor patient prognosis. In conclusion, CGH on cDNA microarrays revealed hundreds of novel genes whose overexpression is attributable to gene amplification. These genes may provide insights to the clonal evolution and progression of breast cancer and highlight promising therapeutic targets.

INTRODUCTION

Gene expression patterns revealed by cDNA microarrays have facilitated classification of cancers into biologically distinct categories, some of which may explain the clinical behavior of the tumors (1–6). Despite this progress in diagnostic classification, the molecular mechanisms underlying gene expression patterns in cancer have remained elusive, and the utility of gene expression profiling in the identification of specific therapeutic targets remains limited.

Accumulation of genetic defects is thought to underlie the clonal evolution of cancer. Identification of the genes that mediate the effects of genetic changes may be important by highlighting transcripts that are actively involved in tumor progression. Such transcripts and their encoded proteins would be ideal targets for anticancer therapies, as demonstrated by the clinical success of new therapies against amplified oncogenes, such as *ERBB2* and *EGFR* (7, 8), in breast cancer and other solid tumors. Besides amplifications of known oncogenes, over 20 recurrent regions of DNA amplification have been mapped in breast cancer by CGH[5] (9, 10). However, these amplicons are often large and poorly defined, and their impact on gene expression remains unknown.

We hypothesized that genome-wide identification of those gene expression changes that are attributable to underlying gene copy number alterations would highlight transcripts that are actively involved in the causation or maintenance of the malignant phenotype. To identify such transcripts, we applied a combination of cDNA and CGH microarrays to: (*a*) determine the global impact that gene copy number variation plays in breast cancer development and progression; and (*b*) identify and characterize those genes whose mRNA expres-

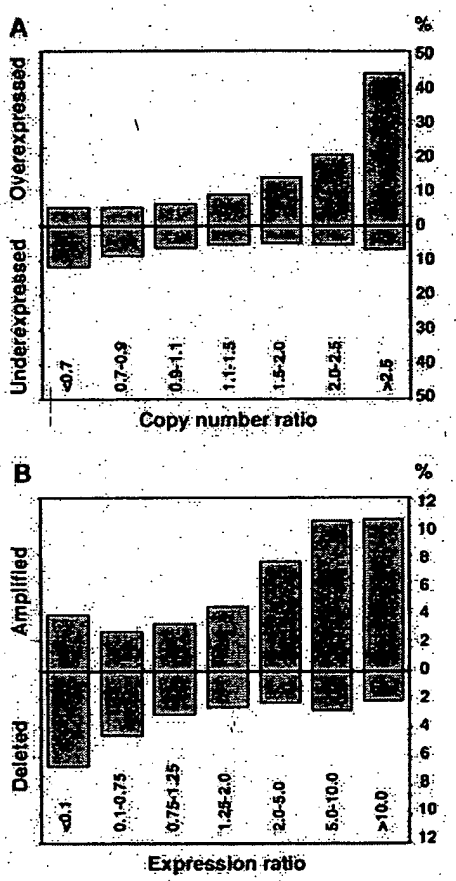


Fig. 1. Impact of gene copy number on global gene expression levels. *A*, percentage of over- and underexpressed genes (*Y axis*) according to copy number ratios (*X axis*). Threshold values used for over- and underexpression were >2.184 (global upper 7% of the cDNA ratios) and <0.4826 (global lower 7% of the expression ratios). *B*, percentage of amplified and deleted genes according to expression ratios. Threshold values for amplification and deletion were >1.5 and <0.7.

Received 5/29/02; accepted 8/28/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Supported in part by the Academy of Finland, Emil Aaltonen Foundation, the Finnish Cancer Society, the Pirkanmaa Cancer Society, the Pirkanmaa Cultural Foundation, the Finnish Breast Cancer Group, the Foundation for the Development of Laboratory Medicine, the Medical Research Fund of the Tampere University Hospital, the Foundation for Commercial and Technical Sciences, and the Swedish Research Council.

[2] Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org).

[3] Contributed equally to this work.

[4] To whom requests for reprints should be addressed, at Laboratory of Cancer Genetics, Institute of Medical Technology, Lenkkeilijankatu 6, FIN-33520 Tampere, Finland. Phone: 358-3247-4125; Fax: 358-3247-4168; E-mail: anne.kallioniemi@uta.fi.

[5] The abbreviations used are: CGH, comparative genomic hybridization; FISH, fluorescence *in situ* hybridization; RT-PCR, reverse transcription-PCR.

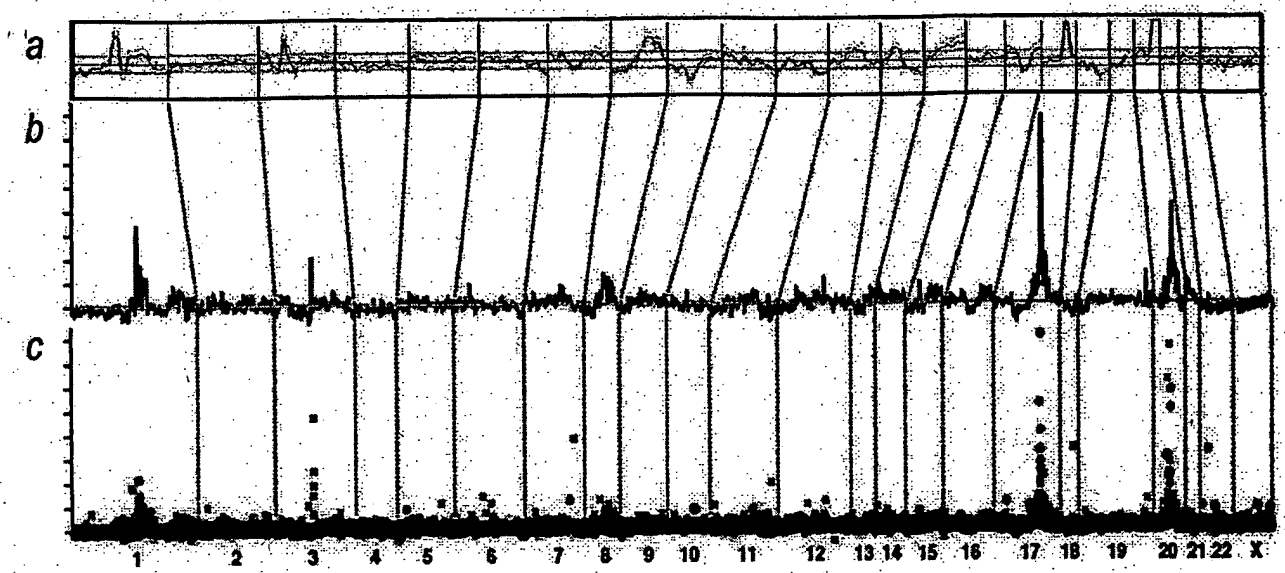


Fig. 2. Genome-wide copy number and expression analysis in the MCF-7 breast cancer cell line. *A*, chromosomal CGH analysis of MCF-7. The copy number ratio profile (*blue line*) across the entire genome from 1p telomere to Xq telomere is shown along with ±1 SD (*orange lines*). The *black horizontal line* indicates a ratio of 1.0; *red line*, a ratio of 0.8; and *green line*, a ratio of 1.2. *B–C*, genome-wide copy number analysis in MCF-7 by CGH on cDNA microarray. The copy number ratios were plotted as a function of the position of the cDNA clones along the human genome. In *B*, individual data points are connected with a line, and a moving median of 10 adjacent clones is shown. *Red horizontal line*, the copy number ratio of 1.0. In *C*, individual data points are labeled by color coding according to cDNA expression ratios. The *bright red dots* indicate the upper 2%, and *dark red dots*, the next 5% of the expression ratios in MCF-7 cells (overexpressed genes); *bright green dots* indicate the lowest 2%, and *dark green dots*, the next 5% of the expression ratios (underexpressed genes); the rest of the observations are shown with *black crosses*. The chromosome numbers are shown at the *bottom* of the figure, and chromosome boundaries are indicated with a *dashed line*.

sion is most significantly associated with amplification of the corresponding genomic template.

MATERIALS AND METHODS

Breast Cancer Cell Lines. Fourteen breast cancer cell lines (BT-20, BT-474, HCC1428, Hs578t, MCF7, MDA-361, MDA-436, MDA-453, MDA-468, SKBR-3, T-47D, UACC812, ZR-75-1, and ZR-75-30) were obtained from the American Type Culture Collection (Manassas, VA). Cells were grown under recommended culture conditions. Genomic DNA and mRNA were isolated using standard protocols.

Copy Number and Expression Analyses by cDNA Microarrays. The preparation and printing of the 13,824 cDNA clones on glass slides were performed as described (11–13). Of these clones, 244 represented uncharacterized expressed sequence tags, and the remainder corresponded to known genes. CGH experiments on cDNA microarrays were done as described (14, 15). Briefly, 20 μg of genomic DNA from breast cancer cell lines and normal human WBCs were digested for 14–18 h with *AluI* and *RsaI* (Life Technologies, Inc., Rockville, MD) and purified by phenol/chloroform extraction. Six μg of digested cell line DNAs were labeled with Cy3-dUTP (Amersham Pharmacia) and normal DNA with Cy5-dUTP (Amersham Pharmacia) using the Bioprime Labeling kit (Life Technologies, Inc.). Hybridization (14, 15) and posthybridization washes (13) were done as described. For the expression analyses, a standard reference (Universal Human Reference RNA; Stratagene, La Jolla, CA) was used in all experiments. Forty μg of reference RNA were labeled with Cy3-dUTP and 3.5 μg of test mRNA with Cy5-dUTP, and the labeled cDNAs were hybridized on microarrays as described (13, 15). For both microarray analyses, a laser confocal scanner (Agilent Technologies, Palo Alto, CA) was used to measure the fluorescence intensities at the target locations using the DEARRAY software (16). After background subtraction, average intensities at each clone in the test hybridization were divided by the average intensity of the corresponding clone in the control hybridization. For the copy number analysis, the ratios were normalized on the basis of the distribution of ratios of all targets on the array and for the expression analysis on the basis of 88 housekeeping genes, which were spotted four times onto the array. Low quality measurements (*i.e.*, copy number data with mean reference intensity <100 fluorescent units, and expression data with both test and reference intensity <100 fluorescent units and/or with spot size <50 units) were excluded from the analysis and were treated as missing values. The distributions of fluorescence ratios were used to define cutpoints for increased/decreased copy number. Genes with CGH ratio >1.43 (representing the upper 5% of the CGH ratios across all experiments) were considered to be amplified, and genes with ratio <0.73 (representing the lower 5%) were considered to be deleted.

Statistical Analysis of CGH and cDNA Microarray Data. To evaluate the influence of copy number alterations on gene expression, we applied the following statistical approach. CGH and cDNA calibrated intensity ratios were log-transformed and normalized using median centering of the values in each cell line. Furthermore, cDNA ratios for each gene across all 14 cell lines were median centered. For each gene, the CGH data were represented by a vector that was labeled 1 for amplification (ratio, >1.43) and 0 for no amplification. Amplification was correlated with gene expression using the signal-to-noise statistics (1). We calculated a weight, w_g, for each gene as follows:

$$w_g = \frac{m_{g1} - m_{g0}}{\sigma_{g1} + \sigma_{g0}}$$

where m_{g1}, σ_{g1} and m_{g0}, σ_{g0} denote the means and SDs for the expression levels for amplified and nonamplified cell lines, respectively. To assess the statistical significance of each weight, we performed 10,000 random permutations of the label vector. The probability that a gene had a larger or equal weight by random permutation than the original weight was denoted by α. A low α (<0.05) indicates a strong association between gene expression and amplification.

Genomic Localization of cDNA Clones and Amplicon Mapping. Each cDNA clone on the microarray was assigned to a Unigene cluster using the Unigene Build 141.[6] A database of genomic sequence alignment information for mRNA sequences was created from the August 2001 freeze of the University of California Santa Cruz's GoldenPath database.[7] The chromosome and bp positions for each cDNA clone were then retrieved by relating these data sets. Amplicons were defined as a CGH copy number ratio >2.0 in at least two adjacent clones in two or more cell lines or a CGH ratio >2.0 in at least three adjacent clones in a single cell line. The amplicon start and end positions were

[6] Internet address: http://research.nhgri.nih.gov/microarray/downloadable_cdna.html.
[7] Internet address: www.genome.ucsc.edu.

Table 1 *Summary of independent amplicons in 14 breast cancer cell lines by CGH microarray*

Location	Start (Mb)	End (Mb)	Size (Mb)
1p13	132.79	132.94	0.2
1q21	173.92	177.25	3.3
1q22	179.28	179.57	0.3
3p14	71.94	74.66	2.7
7p12.1–7p11.2	55.62	60.95	5.3
7q31	125.73	130.96	5.2
7q32	140.01	140.68	0.7
8q21.11–8q21.13	86.45	92.46	6.0
8q21.3	98.45	103.05	4.6
8q23.3–8q24.14	129.88	142.15	12.3
8q24.22	151.21	152.16	1.0
9p13	38.65	39.25	0.6
13q22–q31	77.15	81.38	4.2
16q22	86.70	87.62	0.9
17q11	29.30	30.85	1.6
17q12–q21.2	39.79	42.80	3.0
17q21.32–q21.33	52.47	55.80	3.3
17q22–q23.3	63.81	69.70	5.9
17q23.3–q24.3	69.93	74.99	5.1
19q13	40.63	41.40	0.8
20q11.22	34.59	35.85	1.3
20q13.12	44.00	45.62	1.6
20q13.12–q13.13	46.45	49.43	3.0
20q13.2–q13.32	51.32	59.12	7.8

extended to include neighboring nonamplified clones (ratio, <1.5). The amplicon size determination was partially dependent on local clone density.

FISH. Dual-color interphase FISH to breast cancer cell lines was done as described (17). Bacterial artificial chromosome clone RP11-361K8 was labeled with SpectrumOrange (Vysis, Downers Grove, IL), and SpectrumOrange-labeled probe for *EGFR* was obtained from Vysis. SpectrumGreen-labeled chromosome 7 and 17 centromere probes (Vysis) were used as a reference. A tissue microarray containing 612 formalin-fixed, paraffin-embedded primary breast cancers (17) was applied in FISH analyses as described (18). The use of these specimens was approved by the Ethics Committee of the University of Basel and by the NIH. Specimens containing a 2-fold or higher increase in the number of test probe signals, as compared with corresponding centromere signals, in at least 10% of the tumor cells were considered to be amplified. Survival analysis was performed using the Kaplan-Meier method and the log-rank test.

RT-PCR. The *HOXB7* expression level was determined relative to *GAPDH*. Reverse transcription and PCR amplification were performed using Access RT-PCR System (Promega Corp., Madison, WI) with 10 ng of mRNA as a template. *HOXB7* primers were 5′-GAGCAGAGGGACTCGGACTT-3′ and 5′-GCGTCAGGTAGCGATTGTAG-3′.

RESULTS

Global Effect of Copy Number on Gene Expression. 13,824 arrayed cDNA clones were applied for analysis of gene expression and gene copy number (CGH microarrays) in 14 breast cancer cell lines. The results illustrate a considerable influence of copy number on gene expression patterns. Up to 44% of the highly amplified transcripts (CGH ratio, >2.5) were overexpressed (*i.e.*, belonged to the global upper 7% of expression ratios), compared with only 6% for genes with normal copy number levels (Fig. 1A). Conversely, 10.5% of the transcripts with high-level expression (cDNA ratio, >10) showed increased copy number (Fig. 1*B*). Low-level copy number increases and decreases were also associated with similar, although less dramatic, outcomes on gene expression (Fig. 1).

Identification of Distinct Breast Cancer Amplicons. Base-pair locations obtained for 11,994 cDNAs (86.8%) were used to plot copy number changes as a function of genomic position (Fig. 2, Supplement Fig. A). The average spacing of clones throughout the genome was 267 kb. This high-resolution mapping identified 24 independent breast cancer amplicons, spanning from 0.2 to 12 Mb of DNA (Table 1). Several amplification sites detected previously by chromosomal CGH were validated, with 1q21, 17q12–q21.2, 17q22–q23, 20q13.1, and 20q13.2 regions being most commonly amplified. Furthermore, the boundaries of these amplicons were precisely delineated. In addition, novel amplicons were identified at 9p13 (38.65–39.25 Mb), and 17q21.3 (52.47–55.80 Mb).

Direct Identification of Putative Amplification Target Genes. The cDNA/CGH microarray technique enables the direct correlation of copy number and expression data on a gene-by-gene basis throughout the genome. We directly annotated high-resolution CGH plots with gene expression data using color coding. Fig. 2*C* shows that most of the amplified genes in the MCF-7 breast cancer cell line at 1p13, 17q22–q23, and 20q13 were highly overexpressed. A view of chromosome 7 in the MDA-468 cell line implicates *EGFR* as the most highly overexpressed and amplified gene at 7p11–p12 (Fig. 3*A*). In BT-474, the two known amplicons at 17q12 and 17q22–q23 contained numerous highly overexpressed genes (Fig. 3*B*). In addition, several genes, including the homeobox genes *HOXB2* and *HOXB7*, were highly amplified in a previously undescribed independent amplicon at 17q21.3. *HOXB7* was systematically amplified (as validated by FISH, Fig. 3*B*, *inset*) as well as overexpressed (as verified by RT-PCR, data not shown) in BT-474, UACC812, and ZR-75-30 cells. Furthermore, this novel

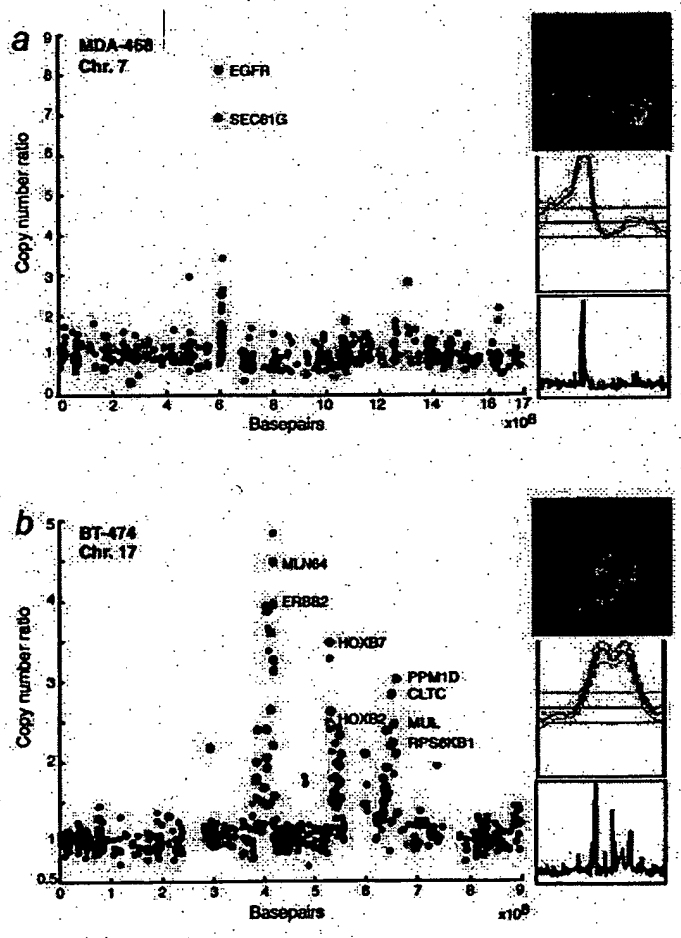


Fig. 3. Annotation of gene expression data on CGH microarray profiles. *A*, genes in the 7p11-p12 amplicon in the MDA-468 cell line are highly expressed (*red dots*) and include the *EGFR* oncogene. *B*, several genes in the 17q12, 17q21.3, and 17q23 amplicons in the BT-474 breast cancer cell line are highly overexpressed (*red*) and include the *HOXB7* gene. The data labels and color coding are as indicated for Fig. 2*C*. *Insets* show chromosomal CGH profiles for the corresponding chromosomes and validation of the increased copy number by interphase FISH using *EGFR* (*red*) and chromosome 7 centromere probe (*green*) to MDA-468 (*A*) and *HOXB7*-specific probe (*red*) and chromosome 17 centromere (*green*) to BT-474 cells (*B*).

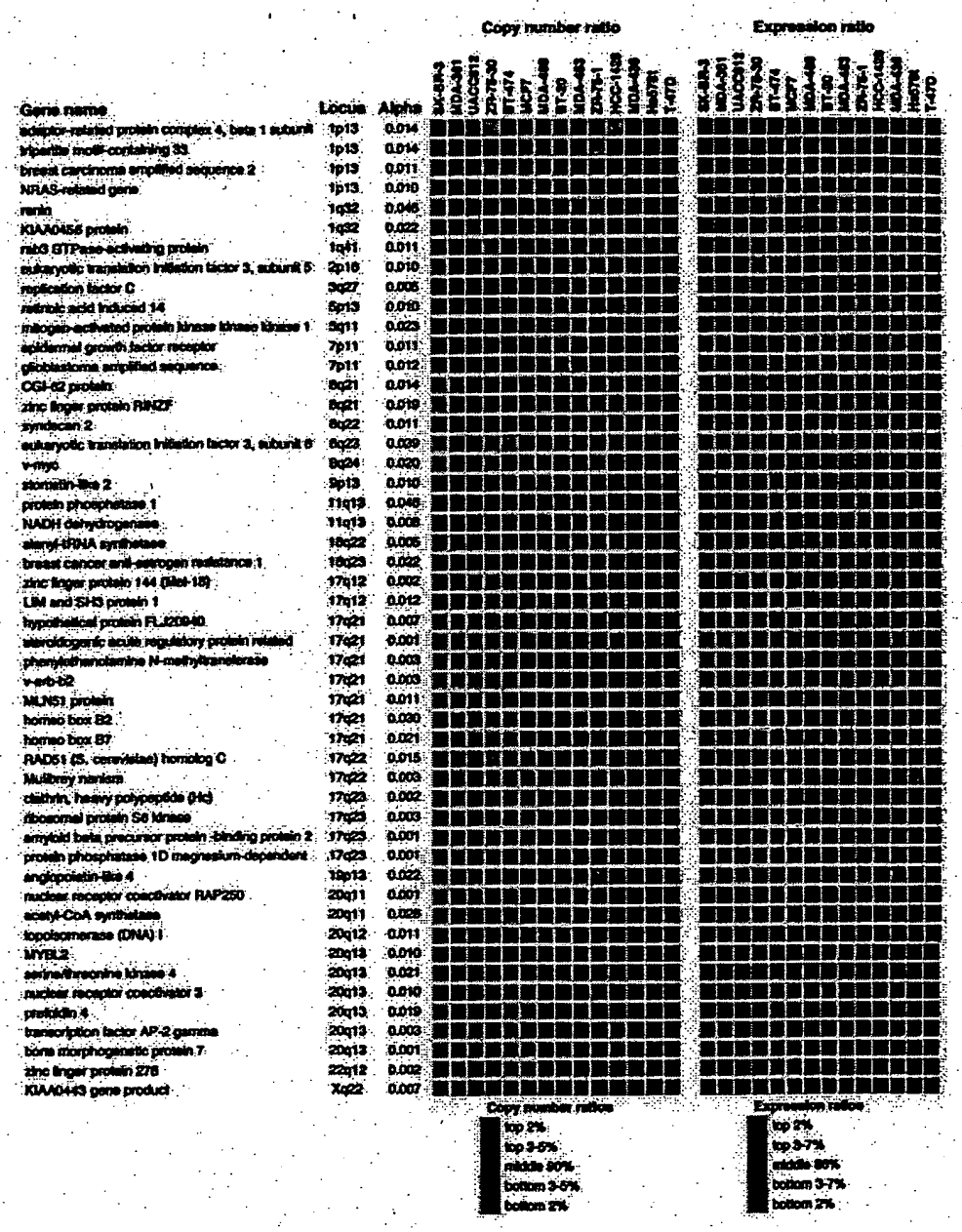


Fig. 4. List of 50 genes with a statistically significant correlation (α value <0.05) between gene copy number and gene expression. Name, chromosomal location, and the α value for each gene are indicated. The genes have been ordered according to their position in the genome. The color maps on the *right* illustrate the copy number and expression ratio patterns in the 14 cell lines. The key to the color code is shown at the *bottom* of the graph. *Gray squares*, missing values. The complete list of 270 genes is shown in supplemental Fig. B.

amplification was validated to be present in 10.2% of 363 primary breast cancers by FISH to a tissue microarray and was associated with poor prognosis of the patients ($P = 0.001$).

Statistical Identification and Characterization of 270 Highly Expressed Genes in Amplicons. Statistical comparison of expression levels of all genes as a function of gene amplification identified 270 genes whose expression was significantly influenced by copy number across all 14 cell lines (Fig. 4, Supplemental Fig. B). According to the gene ontology data,[8] 91 of the 270 genes represented hypothetical proteins or genes with no functional annotation, whereas 179 had associated functional information available. Of these, 151 (84%) are implicated in apoptosis, cell proliferation, signal transduction, and transcription, whereas 28 (16%) had functional annotations that could not be directly linked with cancer.

DISCUSSION

The importance of recurrent gene and chromosome copy number changes in the development and progression of solid tumors has been characterized in >1000 publications applying CGH[9] (9, 10), as well as in a large number of other molecular cytogenetic, cytogenetic, and molecular genetic studies. The effects of these somatic genetic changes on gene expression levels have remained largely unknown, although a few studies have explored gene expression changes occurring in specific amplicons (15, 19–21). Here, we applied genome-wide cDNA microarrays to identify transcripts whose expression changes were attributable to underlying gene copy number alterations in breast cancer.

The overall impact of copy number on gene expression patterns was substantial with the most dramatic effects seen in the case of high-

[8] Internet address: http://www.geneontology.org/.

[9] Internet address: http://www.ncbi.nlm.nih.gov/entrez.

level copy number increase. Low-level copy number gains and losses also had a significant influence on expression levels of genes in the regions affected, but these effects were more subtle on a gene-by-gene basis than those of high-level amplifications. However, the impact of low-level gains on the dysregulation of gene expression patterns in cancer may be equally important if not more important than that of high-level amplifications. Aneuploidy and low-level gains and losses of chromosomal arms represent the most common types of genetic alterations in breast and other cancers and, therefore, have an influence on many genes. Our results in breast cancer extend the recent studies on the impact of aneuploidy on global gene expression patterns in yeast cells, acute myeloid leukemia, and a prostate cancer model system (22–24).

The CGH microarray analysis identified 24 independent breast cancer amplicons. We defined the precise boundaries for many amplicons detected previously by chromosomal CGH (9, 10, 25, 26) and also discovered novel amplicons that had not been detected previously, presumably because of their small size (only 1–2 Mb) or close proximity to other larger amplicons. One of these novel amplicons involved the homeobox gene region at 17q21.3 and led to the overexpression of the *HOXB7* and *HOXB2* genes. The homeodomain transcription factors are known to be key regulators of embryonic development and have been occasionally reported to undergo aberrant expression in cancer (27, 28). *HOXB7* transfection induced cell proliferation in melanoma, breast, and ovarian cancer cells and increased tumorigenicity and angiogenesis in breast cancer (29–32). The present results imply that gene amplification may be a prominent mechanism for overexpressing *HOXB7* in breast cancer and suggest that *HOXB7* contributes to tumor progression and confers an aggressive disease phenotype in breast cancer. This view is supported by our finding of amplification of *HOXB7* in 10% of 363 primary breast cancers, as well as an association of amplification with poor prognosis of the patients.

We carried out a systematic search to identify genes whose expression levels across all 14 cell lines were attributable to amplification status. Statistical analysis revealed 270 such genes (representing ~2% of all genes on the array), including not only previously described amplified genes, such as *HER-2*, *MYC*, *EGFR*, ribosomal protein s6 kinase, and *AIB3*, but also numerous novel genes such as *NRAS-related gene* (1p13), *syndecan-2* (8q22), and *bone morphogenic protein* (20q13.1), whose activation by amplification may similarly promote breast cancer progression. Most of the 270 genes have not been implicated previously in breast cancer development and suggest novel pathogenetic mechanisms. Although we would not expect all of them to be causally involved, it is intriguing that 84% of the genes with associated functional information were implicated in apoptosis, cell proliferation, signal transduction, transcription, or other cellular processes that could directly imply a possible role in cancer progression. Therefore, a detailed characterization of these genes may provide biological insights to breast cancer progression and might lead to the development of novel therapeutic strategies.

In summary, we demonstrate application of cDNA microarrays to the analysis of both copy number and expression levels of over 12,000 transcripts throughout the breast cancer genome, roughly once every 267 kb. This analysis provided: (*a*) evidence of a prominent global influence of copy number changes on gene expression levels; (*b*) a high-resolution map of 24 independent amplicons in breast cancer; and (*c*) identification of a set of 270 genes, the overexpression of which was statistically attributable to gene amplification. Characterization of a novel amplicon at 17q21.3 implicated amplification and overexpression of the *HOXB7* gene in breast cancer, including a clinical association between *HOXB7* amplification and poor patient prognosis. Overall, our results illustrate how the identification of genes activated by gene amplification provides a powerful approach to highlight genes with an important role in cancer as well as to prioritize and validate putative targets for therapy development.

REFERENCES

1. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science (Wash. DC), *286:* 531–537, 1999.
2. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature (Lond.), *403:* 503–511, 2000.
3. Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., *et al.* Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature (Lond.), *406:* 536–540, 2000.
4. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., *et al.* Molecular portraits of human breast tumours. Nature (Lond.), *406:* 747–752, 2000.
5. Dhanasekaran, S. M., Barrette, T. R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K. J., Rubin, M. A., and Chinnaiyan, A. M. Delineation of prognostic biomarkers in prostate cancer. Nature (Lond.), *412:* 822–826, 2001.
6. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. USA, *98:* 10869–10874, 2001.
7. Ross, J. S., and Fletcher, J. A. The *HER-2/neu* oncogene: prognostic factor, predictive factor and target for therapy. Semin. Cancer Biol., *9:* 125–138, 1999.
8. Arteaga, C. L. The epidermal growth factor receptor: from mutant oncogene in nonhuman cancers to therapeutic target in human neoplasia. J. Clin. Oncol., *19:* 32–40, 2001.
9. Knuutila, S., Bjorkqvist, A. M., Autio, K., Tarkkanen, M., Wolf, M., Monni, O., Szymanska, J., Larramendy, M. L., Tapper, J., Pere, H., El-Rifai, W., *et al.* DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. Am. J. Pathol., *152:* 1107–1123, 1998.
10. Knuutila S., Autio K., and Aalto Y. Online access to CGH data of DNA sequence copy number changes. Am. J. Pathol., *157:* 689, 2000.
11. DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat. Genet., *14:* 457–460, 1996.
12. Shalon, D., Smith, S. J., and Brown, P. O. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res., *6:* 639–645, 1996.
13. Mousses, S., Bittner, M. L., Chen, Y., Dougherty, E. R., Baxevanis, A., Meltzer, P. S., and Trent, J. M. Gene expression analysis by cDNA microarrays. *In*: F. J. Livesey and S. P. Hunt (eds.), Functional Genomics, pp. 113–137. Oxford: Oxford University Press, 2000.
14. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat. Genet., *23:* 41–46, 1999.
15. Monni, O., Bärlund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M. L., and Kallioniemi, A. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. Proc. Natl. Acad. Sci. USA, *98:* 5711–5716, 2001.
16. Chen, Y., Dougherty, E. R., and Bittner, M. L. Ratio-based decisions and the quantitative analysis of cDNA microarray images. J. Biomed. Optics, *2:* 364–374, 1997.
17. Bärlund, M., Forozan, F., Kononen, J., Bubendorf, L., Chen, Y., Bittner, M. L., Torhorst, J., Haas, P., Bucher, C., Sauter, G., *et al.* Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. J. Natl. Cancer Inst., *92:* 1252–1259, 2000.
18. Andersen, C. L., Hostetter, G., Grigoryan, A., Sauter, G., and Kallioniemi, A. Improved procedure for fluorescence *in situ* hybridization on tissue microarrays. Cytometry, *45:* 83–86, 2001.
19. Kauraniemi, P., Bärlund, M., Monni, O., and Kallioniemi, A. New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. Cancer Res., *61:* 8235–8240, 2001.
20. Clark, J., Edwards, S., John, M., Flohr, P., Gordon, T., Maillard, K., Giddings, I., Brown, C., Bagherzadeh, A., Campbell, C., Shipley, J., Wooster, R., and Cooper, C. S. Identification of amplified and expressed genes in breast cancer by comparative hybridization onto microarrays of randomly selected cDNA clones. Genes Chromosomes Cancer, *34:* 104–114, 2002.
21. Varis, A., Wolf, M., Monni, O., Vakkari, M. L., Kokkola, A., Moskaluk, C., Frierson, H., Powell, S. M., Knuutila, S., Kallioniemi, A., and El-Rifai, W. Targets of gene amplification and overexpression at 17q in gastric cancer. Cancer Res., *62:* 2625–2629, 2002.
22. Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., Friend, S. H., and Marton M. J.

Widespread aneuploidy revealed by DNA microarray expression profiling. Nat. Genet., *25:* 333–337, 2000.

23. Virtaneva, K., Wright, F. A., Tanner, S. M., Yuan, B., Lemon, W. J., Caligiuri, M. A., Bloomfield, C. D., de La Chapelle, A., and Krahe, R. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. Proc. Natl. Acad. Sci. USA, *98:* 1124–1129, 2001.
24. Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., Linehan, W. M., Cunha, G. R., and Ried, T. The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. Cancer Res., *61:* 8143–8149, 2001.
25. Bärlund, M., Tirkkonen, M., Forozan, F., Tanner, M. M., Kallioniemi, O. P., and Kallioniemi, A. Increased copy number at 17q22–q24 by CGH in breast cancer is due to high-level amplification of two separate regions. Genes Chromosomes Cancer, *20:* 372–376, 1997.
26. Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Isola, J., Kuukasjärvi, T., Collins, C., Kowbel, D., Guan, X. Y., Trent, J., Gray, J. W., Meltzer, P., and Kallioniemi O. P. Independent amplification and frequent co-amplification of three nonsyntenić regions on the long arm of chromosome 20 in human breast cancer. Cancer Res., *56:* 3441–3445, 1996.
27. Cillo, C., Faiella, A., Cantile, M., and Boncinelli, E. Homeobox genes and cancer. Exp. Cell Res., *248:* 1–9, 1999.
28. Cillo, C., Cantile, M., Faiella, A., and Boncinelli, E. Homeobox genes in normal and malignant cells. J. Cell. Physiol., *188:* 161–169, 2001.
29. Care, A., Silvani, A., Meccia, E., Mattia, G., Stoppacciaro, A., Parmiani, G., Peschle, C., and Colombo, M. P. HOXB7 constitutively activates basic fibroblast growth factor in melanomas. Mol. Cell. Biol., *16:* 4842–4851, 1996.
30. Care, A., Silvani, A., Meccia, E., Mattia, G., Peschle, C., and Colombo, M. P. Transduction of the SkBr3 breast carcinoma cell line with the *HOXB7* gene induces bFGF expression, increases cell proliferation and reduces growth factor dependence. Oncogene, *16:* 3285–3289, 1998.
31. Care, A., Felicetti, F., Meccia, E., Bottero, L., Parenza, M., Stoppacciaro, A., Peschle, C., and Colombo, M. P. HOXB7: a key factor for tumor-associated angiogenic switch. Cancer Res., *61:* 6532–6539, 2001.
32. Naora, H., Yang, Y. Q., Montz, F. J., Seidman, J. D., Kurman, R. J., and Roden, R. B. A serologically identified tumor antigen encoded by a homeobox gene promotes growth of ovarian epithelial cells. Proc. Natl. Acad. Sci. USA, *98:* 4060–4065, 2001.

Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors

Jonathan R. Pollack*†‡, Therese Sørlie§, Charles M. Perou¶, Christian A. Rees‖**, Stefanie S. Jeffrey††, Per E. Lonning‡‡, Robert Tibshirani§§, David Botstein‖, Anne-Lise Børresen-Dale§, and Patrick O. Brown†¶¶

Departments of *Pathology, ‖Genetics, ††Surgery, §§Health Research and Policy, and ¶¶Biochemistry, and †Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305; §Department of Genetics, Norwegian Radium Hospital, Montebello, N-0310 Oslo, Norway; ‡‡Department of Medicine (Oncology), Haukeland University Hospital, N-5021 Bergen, Norway; and ¶Department of Genetics and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599

Contributed by Patrick O. Brown, August 6, 2002

Genomic DNA copy number alterations are key genetic events in the development and progression of human cancers. Here we report a genome-wide microarray comparative genomic hybridization (array CGH) analysis of DNA copy number variation in a series of primary human breast tumors. We have profiled DNA copy number alteration across 6,691 mapped human genes, in 44 predominantly advanced, primary breast tumors and 10 breast cancer cell lines. While the overall patterns of DNA amplification and deletion corroborate previous cytogenetic studies, the high-resolution (gene-by-gene) mapping of amplicon boundaries and the quantitative analysis of amplicon shape provide significant improvement in the localization of candidate oncogenes. Parallel microarray measurements of mRNA levels reveal the remarkable degree to which variation in gene copy number contributes to variation in gene expression in tumor cells. Specifically, we find that 62% of highly amplified genes show moderately or highly elevated expression, that DNA copy number influences gene expression across a wide range of DNA copy number alterations (deletion, low-, mid- and high-level amplification), that on average, a 2-fold change in DNA copy number is associated with a corresponding 1.5-fold change in mRNA levels, and that overall, at least 12% of all the variation in gene expression among the breast tumors is directly attributable to underlying variation in gene copy number. These findings provide evidence that widespread DNA copy number alteration can lead directly to global deregulation of gene expression, which may contribute to the development or progression of cancer.

Conventional cytogenetic techniques, including comparative genomic hybridization (CGH) (1), have led to the identification of a number of recurrent regions of DNA copy number alteration in breast cancer cell lines and tumors (2–4). While some of these regions contain known or candidate oncogenes [e.g., FGFR1 (8p11), MYC (8q24), CCND1 (11q13), ERBB2 (17q12), and ZNF217 (20q13)] and tumor suppressor genes [RB1 (13q14) and TP53 (17p13)], the relevant gene(s) within other regions (e.g., gain of 1q, 8q22, and 17q22–24, and loss of 8p) remain to be identified. A high-resolution genome-wide map, delineating the boundaries of DNA copy number alterations in tumors, should facilitate the localization and identification of oncogenes and tumor suppressor genes in breast cancer. In this study, we have created such a map, using array-based CGH (5–7) to profile DNA copy number alteration in a series of breast cancer cell lines and primary tumors.

An unresolved question is the extent to which the widespread DNA copy number changes that we and others have identified in breast tumors alter expression of genes within involved regions. Because we had measured mRNA levels in parallel in the same samples (8), using the same DNA microarrays, we had an opportunity to explore on a genomic scale the relationship between DNA copy number changes and gene expression. From this analysis, we have identified a significant impact of widespread DNA copy number alteration on the transcriptional programs of breast tumors.

Materials and Methods

Tumors and Cell Lines. Primary breast tumors were predominantly large (>3 cm), intermediate-grade, infiltrating ductal carcinomas, with more than 50% being lymph node positive. The fraction of tumor cells within specimens averaged at least 50%. Details of individual tumors have been published (8, 9), and are summarized in Table 1, which is published as supporting information on the PNAS web site, www.pnas.org. Breast cancer cell lines were obtained from the American Type Culture Collection. Genomic DNA was isolated either using Qiagen genomic DNA columns, or by phenol/chloroform extraction followed by ethanol precipitation.

DNA Labeling and Microarray Hybridizations. Genomic DNA labeling and hybridizations were performed essentially as described in Pollack *et al.* (7), with slight modifications. Two micrograms of DNA was labeled in a total volume of 50 microliters and the volumes of all reagents were adjusted accordingly. "Test" DNA (from tumors and cell lines) was fluorescently labeled (Cy5) and hybridized to a human cDNA microarray containing 6,691 different mapped human genes (i.e., UniGene clusters). The "reference" (labeled with Cy3) for each hybridization was normal female leukocyte DNA from a single donor. The fabrication of cDNA microarrays and the labeling and hybridization of mRNA samples have been described (8).

Data Analysis and Map Positions. Hybridized arrays were scanned on a GenePix scanner (Axon Instruments, Foster City, CA), and fluorescence ratios (test/reference) calculated using SCANALYZE software (available at http://rana.lbl.gov). Fluorescence ratios were normalized for each array by setting the average log fluorescence ratio for all array elements equal to 0. Measurements with fluorescence intensities more than 20% above background were considered reliable. DNA copy number profiles that deviated significantly from background ratios measured in normal genomic DNA control hybridizations were interpreted as evidence of real DNA copy number alteration (see *Estimating Significance of Altered Fluorescence Ratios* in the supporting information). When indicated, DNA copy number profiles are displayed as a moving average (symmetric 5-nearest neighbors). Map positions for arrayed human cDNAs were assigned by

Abbreviation: CGH, comparative genomic hybridization.

†To whom reprint requests should be addressed at: Department of Pathology, Stanford University School of Medicine, CCSR Building, Room 3245A, 269 Campus Drive, Stanford, CA 94305-5176. E-mail: pollack1@stanford.edu.

**Present address: Zyomyx Inc., Hayward, CA 94545.

GENETICS

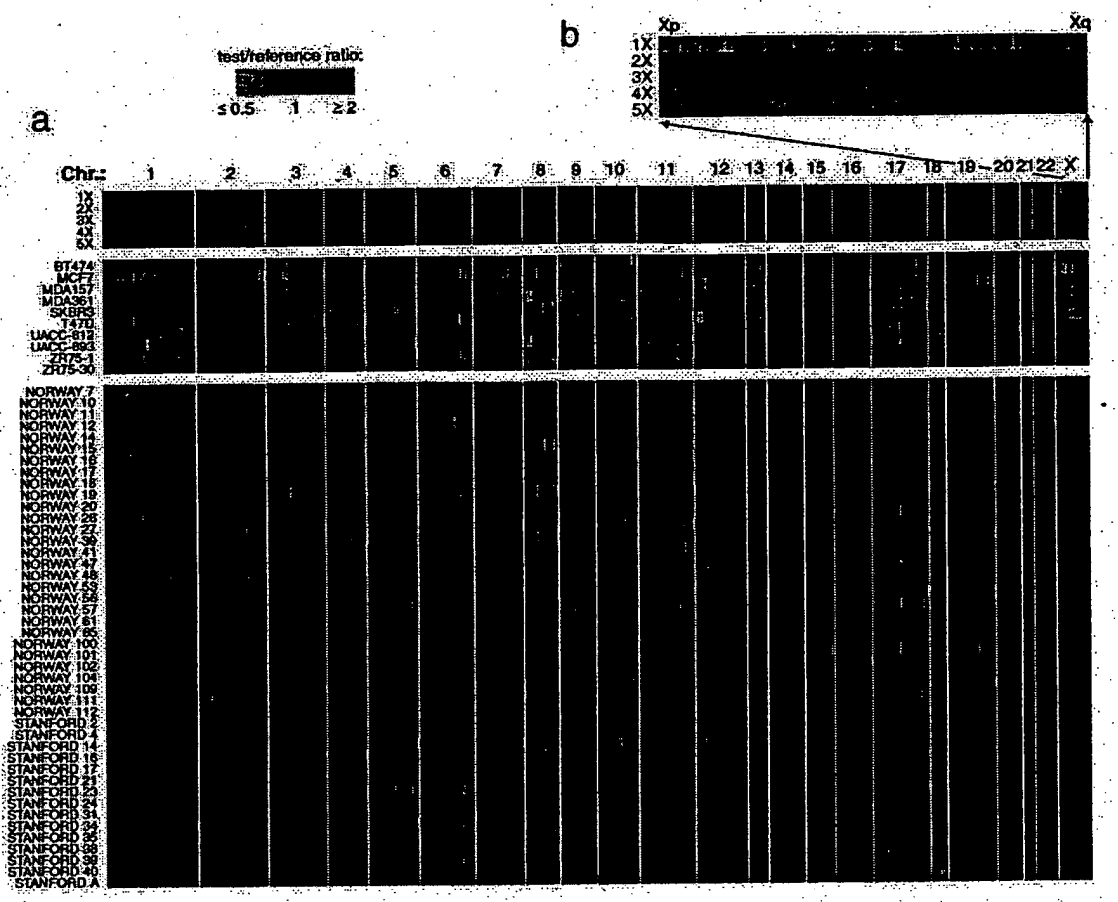


Fig. 1. Genome-wide measurement of DNA copy number alteration by array CGH. (*a*) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Each row represents a different cell line or tumor, and each column represents one of 6,691 different mapped human genes present on the microarray, ordered by genome map position from 1pter through Xqter. Moving average (symmetric 5-nearest neighbors) fluorescence ratios (test/reference) are depicted using a $\log_2$-based pseudocolor scale (indicated), such that red luminescence reflects fold-amplification, green luminescence reflects fold-deletion, and black indicates no change (gray indicates poorly measured data). (*b*) Enlarged view of DNA copy number profiles across the X chromosome, shown for cell lines containing different numbers of X chromosomes.

identifying the starting position of the best and longest match of any DNA sequence represented in the corresponding UniGene cluster (10) against the "Golden Path" genome assembly (http://genome.ucsc.edu/; Oct 7, 2000 Freeze). For UniGene clusters represented by multiple arrayed elements, mean fluorescence ratios (for all elements representing the same UniGene cluster) are reported. For mRNA measurements, fluorescence ratios are "mean-centered" (i.e., reported relative to the mean ratio across the 44 tumor samples). The data set described here can be accessed in its entirety in the supporting information.

Results

We performed CGH on 44 predominantly locally advanced, primary breast tumors and 10 breast cancer cell lines, using cDNA microarrays containing 6,691 different mapped human genes (Fig. 1*a*; also see *Materials and Methods* for details of microarray hybridizations). To take full advantage of the improved spatial resolution of array CGH, we ordered (fluorescence ratios for) the 6,691 cDNAs according to the "Golden Path" (http://genome.ucsc.edu/) genome assembly of the draft human genome sequences (11). In so doing, arrayed cDNAs not only themselves represent genes of potential interest (e.g., candidate oncogenes within amplicons), but also provide precise genetic landmarks for chromosomal regions of amplification and deletion. Parallel analysis of DNA from cell lines containing different numbers of X chromosomes (Fig. 1*b*), as we did before (7), demonstrated the sensitivity of our method to detect single-copy loss (45, XO), and 1.5- (47,XXX), 2- (48,XXXX), or 2.5-fold (49,XXXXX) gains (also see Fig. 5, which is published as supporting information on the PNAS web site). Fluorescence ratios were linearly proportional to copy number ratios, which were slightly underestimated, in agreement with previous observations (7). Numerous DNA copy number alterations were evident in both the breast cancer cell lines and primary tumors (Fig. 1*a*), detected in the tumors despite the presence of euploid non-tumor cell types; the magnitudes of the observed changes were generally lower in the tumor samples. DNA copy-number alterations were found in every cancer cell line and tumor, and on every human chromosome in at least one sample. Recurrent regions of DNA copy number gain and loss were readily identifiable. For example, gains within 1q, 8q, 17q, and 20q were observed in a high proportion of breast cancer cell lines/tumors (90%/69%, 100%/47%, 100%/60%, and 90%/44%, respectively), as were losses within 1p, 3p, 8p, and 13q (80%/24%, 80%/22%, 80%/22%, and 70%/18%, respectively), consistent with published cytogenetic studies (refs. 2–4; a complete listing of gains/losses is provided in Tables 2 and 3, which are published as supporting information on the PNAS web site). The total

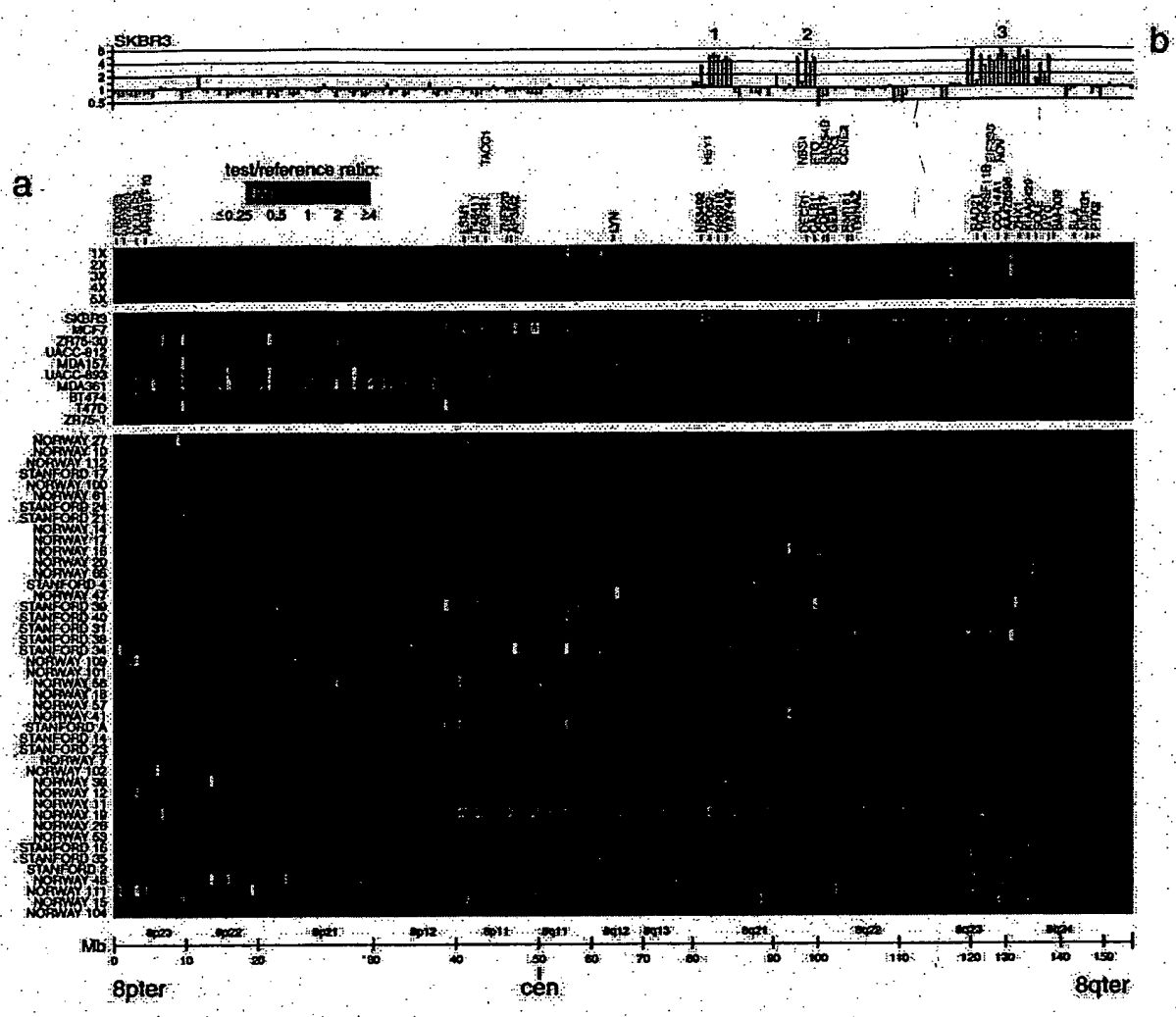


Fig. 2. DNA copy number alteration across chromosome 8 by array CGH. (*a*) DNA copy number profiles are illustrated for cell lines containing different numbers of X chromosomes, for breast cancer cell lines, and for breast tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering to highlight recurrent copy number changes. The 241 genes present on the microarrays and mapping to chromosome 8 are ordered by position along the chromosome. Fluorescence ratios (test/reference) are depicted by a $\log_2$ pseudocolor scale (indicated). Selected genes are indicated with color-coded text (red, increased; green, decreased; black, no change; gray, not well measured) to reflect correspondingly altered mRNA levels (observed in the majority of the subset of samples displaying the DNA copy number change). The map positions for genes of interest that are not represented on the microarray are indicated in the row above those genes represented on the array. (*b*) Graphical display of DNA copy number profile for breast cancer cell line SKBR3. Fluorescence ratios (tumor/normal) are plotted on a $\log_2$ scale for chromosome 8 genes, ordered along the chromosome.

number of genomic alterations (gains and losses) was found to be significantly higher in breast tumors that were high grade ($P = 0.008$), consistent with published CGH data (3), estrogen receptor negative ($P = 0.04$), and harboring TP53 mutations ($P = 0.0006$) (see Table 4, which is published as supporting information on the PNAS web site).

The improved spatial resolution of our array CGH analysis is illustrated for chromosome 8, which displayed extensive DNA copy number alteration in our series. A detailed view of the variation in the copy number of 241 genes mapping to chromosome 8 revealed multiple regions of recurrent amplification; each of these potentially harbors a different known or previously uncharacterized oncogene (Fig. 2*a*). The complexity of amplicon structure is most easily appreciated in the breast cancer cell line SKBR3. Although a conventional CGH analysis of 8q in SKBR3 identified only two distinct regions of amplification (12), we observed three distinct regions of high-level amplification (labeled 1–3 in Fig. 2*b*). For each of these regions we can define the boundaries of the interval recurrently amplified in the tumors we examined; in each case, known or plausible candidate oncogenes can be identified (a description of these regions, as well as the recurrently amplified regions on chromosomes 17 and 20, can be found in Figs. 6 and 7, which are published as supporting information on the PNAS web site).

For a subset of breast cancer cell lines and tumors (4 and 37, respectively), and a subset of arrayed genes (6,095), mRNA levels were quantitatively measured in parallel by using cDNA microarrays (8). The parallel assessment of mRNA levels is useful in the interpretation of DNA copy number changes. For example, the highly amplified genes that are also highly expressed are the strongest candidate oncogenes within an amplicon. Perhaps more significantly, our parallel analysis of DNA copy number changes and mRNA levels provides us the opportunity to assess the global impact of widespread DNA copy number alteration on gene expression in tumor cells.

A strong influence of DNA copy number on gene expression is evident in an examination of the pseudocolor representations

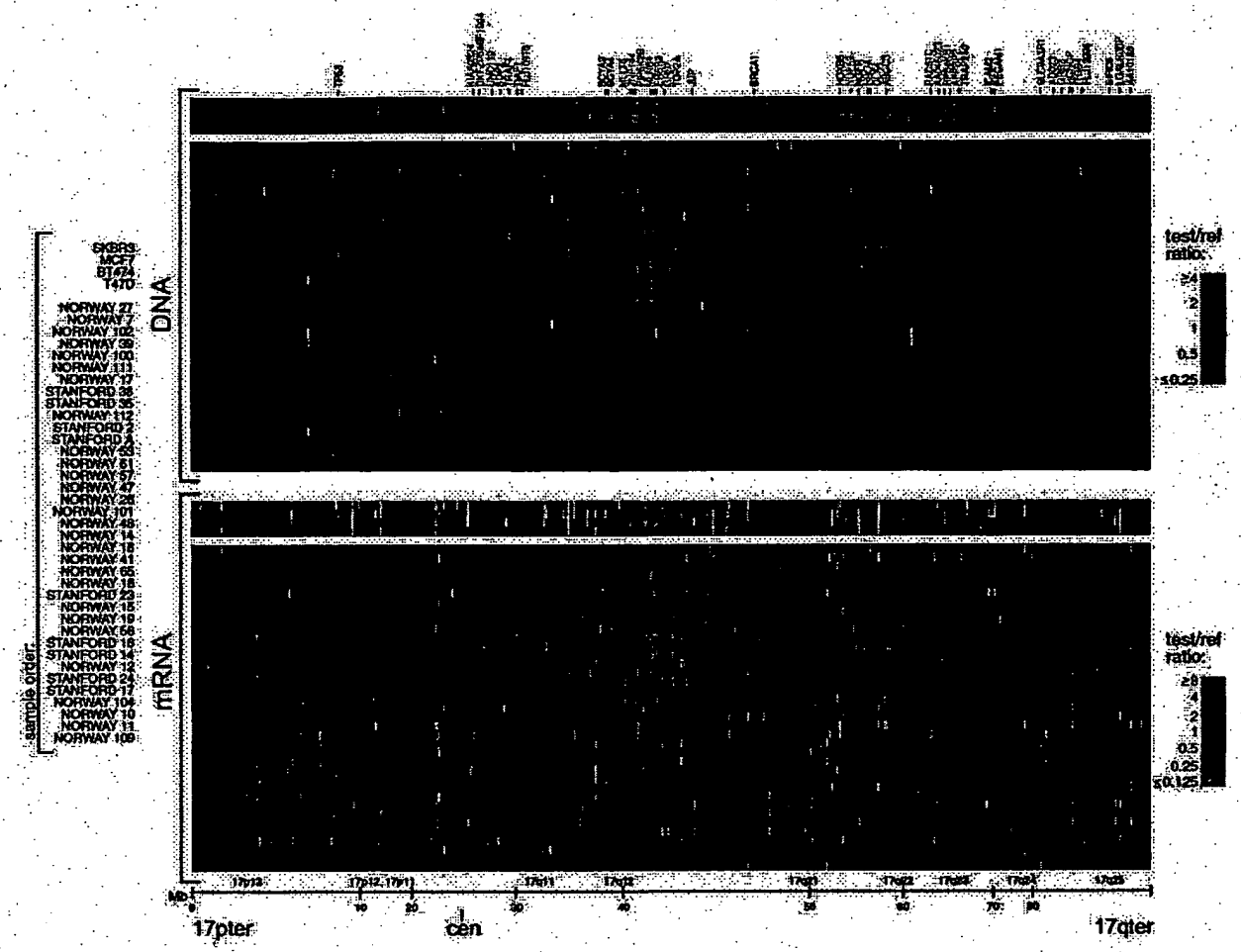


Fig. 3. Concordance between DNA copy number and gene expression across chromosome 17. DNA copy number alteration (*Upper*) and mRNA levels (*Lower*) are illustrated for breast cancer cell lines and tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering (*Upper*), and the identical sample order is maintained (*Lower*). The 354 genes present on the microarrays and mapping to chromosome 17, and for which both DNA copy number and mRNA levels were determined, are ordered by position along the chromosome; selected genes are indicated in color-coded text (see Fig. 2 legend). Fluorescence ratios (test/reference) are depicted by separate $\log_2$ pseudocolor scales (indicated).

of DNA copy number and mRNA levels for genes on chromosome 17 (Fig. 3). The overall patterns of gene amplification and elevated gene expression are quite concordant; i.e., a significant fraction of highly amplified genes appear to be correspondingly highly expressed. The concordance between high-level amplification and increased gene expression is not restricted to chromosome 17. Genome-wide, of 117 high-level DNA amplifications (fluorescence ratios >4, and representing 91 different genes), 62% (representing 54 different genes; see Table 5, which is published as supporting information on the PNAS web site) are found associated with at least moderately elevated mRNA levels (mean-centered fluorescence ratios >2), and 42% (representing 36 different genes) are found associated with comparably highly elevated mRNA levels (mean-centered fluorescence ratios >4).

To determine the extent to which DNA deletion and lower-level amplification (in addition to high-level amplification) are also associated with corresponding alterations in mRNA levels, we performed three separate analyses on the complete data set (4 cell lines and 37 tumors, across 6,095 genes). First, we determined the average mRNA levels for each of five classes of genes, representing DNA deletion, no change, and low-, medium-, and high-level amplification (Fig. 4*a*). For both the breast cancer cell lines and tumors, average mRNA levels tracked with DNA copy number across all five classes, in a statistically significant fashion (P values for pair-wise Student's t tests comparing adjacent classes: cell lines, 4×10^{-49}, 1×10^{-49}, 5×10^{-5}, 1×10^{-2}; tumors, 1×10^{-43}, 1×10^{-214}, 5×10^{-41}, 1×10^{-4}). A linear regression of the average log(DNA copy number), for each class, against average log(mRNA level) demonstrated that on average, a 2-fold change in DNA copy number was accompanied by 1.4- and 1.5-fold changes in mRNA level for the breast cancer cell lines and tumors, respectively (Fig. 4*a*, regression line not shown). Second, we characterized the distribution of the 6,095 correlations between DNA copy number and mRNA level, each across the 37 tumor samples (Fig. 4*b*). The distribution of correlations forms a normal-shaped curve, but with the peak markedly shifted in the positive direction from zero. This shift is statistically significant, as evidenced in a plot of observed vs. expected correlations (Fig. 4*c*), and reflects a pervasive global influence of DNA copy number alterations on gene expression. Notably, the highest correlations between DNA copy number and mRNA level (the right tail of the distribution in Fig. 4*b*) comprise both amplified and deleted genes (data not shown). Third, we used a linear regression model to estimate the fraction of all variation measured in mRNA levels among the 37

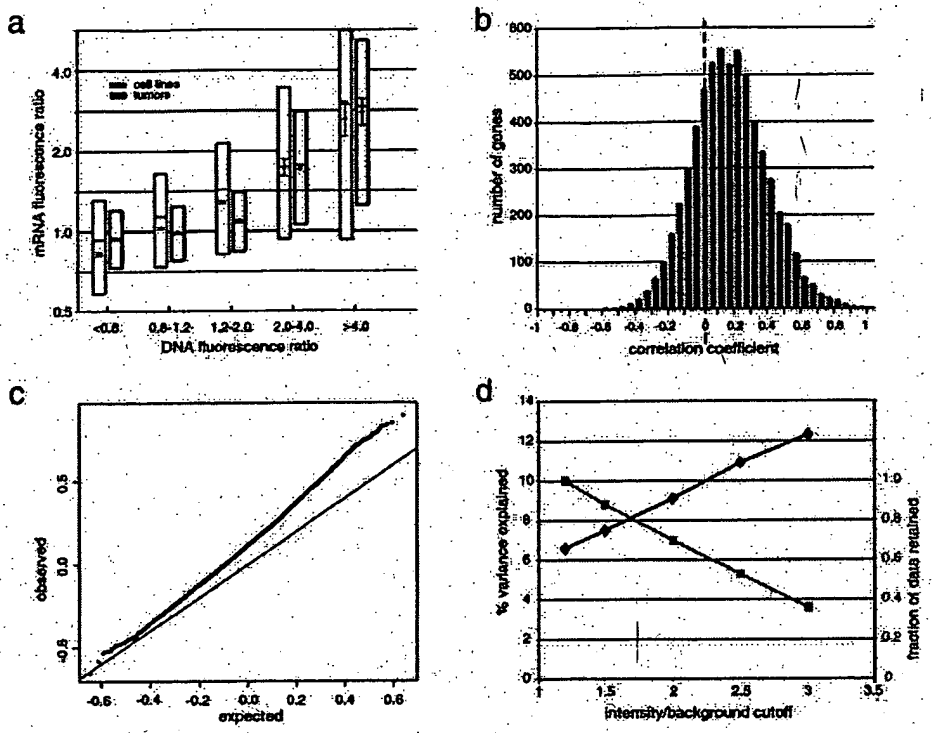


Fig. 4. Genome-wide influence of DNA copy number alterations on mRNA levels. (*a*) For breast cancer cell lines (gray) and tumor samples (black), both mean-centered mRNA fluorescence ratio (log_2 scale) quartiles (box plots indicate 25th, 50th, and 75th percentile) and averages (diamonds; *Y*-value error bars indicate standard errors of the mean) are plotted for each of five classes of genes, representing DNA deletion (tumor/normal ratio < 0.8), no change (0.8–1.2), low- (1.2–2), medium- (2–4), and high-level (>4) amplification. *P* values for pair-wise Student's *t* tests, comparing averages between adjacent classes (moving left to right), are 4×10^{-49}, 1×10^{-49}, 5×10^{-5}, 1×10^{-2} (cell lines), and 1×10^{-43}, 1×10^{-214}, 5×10^{-41}, 1×10^{-4} (tumors). (*b*) Distribution of correlations between DNA copy number and mRNA levels, for 6,095 different human genes across 37 breast tumor samples. (*c*) Plot of observed versus expected correlation coefficients. The expected values were obtained by randomization of the sample labels in the DNA copy number data set. The line of unity is indicated. (*d*) Percent variance in gene expression (among tumors) directly explained by variation in gene copy number. Percent variance explained (black line) and fraction of data retained (gray line) are plotted for different fluorescence intensity/background (a rough surrogate for signal/noise) cutoff values. Fraction of data retained is relative to the 1.2 intensity/background cutoff. Details of the linear regression model used to estimate the fraction of variation in gene expression attributable to underlying DNA copy number alteration can be found in the supporting information (see *Estimating the Fraction of Variation in Gene Expression Attributable to Underlying DNA Copy Number Alteration*).

tumors that could be attributed to underlying variation in DNA copy number. From this analysis, we estimate that, overall, about 7% of all of the observed variation in mRNA levels can be explained directly by variation in copy number of the altered genes (Fig. 4*d*). We can reduce the effects of experimental measurement error on this estimate by using only that fraction of the data most reliably measured (fluorescence intensity/background >3); using that data, our estimate of the percent variation in mRNA levels directly attributed to variation in gene copy number increases to 12% (Fig. 4*d*). This still undoubtedly represents a significant underestimate, as the observed variation in global gene expression is affected not only by true variation in the expression programs of the tumor cells themselves, but also by the variable presence of non-tumor cell types within clinical samples.

Discussion

This genome-wide, array CGH analysis of DNA copy number alteration in a series of human breast tumors demonstrates the usefulness of defining amplicon boundaries at high resolution (gene-by-gene), and quantitatively measuring amplicon shape, to assist in locating and identifying candidate oncogenes. By analyzing mRNA levels in parallel, we have also discovered that changes in DNA copy number have a large, pervasive, direct effect on global gene expression patterns in both breast cancer cell lines and tumors. Although the DNA microarrays used in our analysis may display a bias toward characterized and/or highly expressed genes, because we are examining such a large fraction of the genome (approximately 20% of all human genes), and because, as detailed above, we are likely underestimating the contribution of DNA copy number changes to altered gene expression, we believe our findings are likely to be generalizable (but would nevertheless still be remarkable if only applicable to this set of ~6,100 genes).

In budding yeast, aneuploidy has been shown to result in chromosome-wide gene expression biases (13). Two recent studies have begun to examine the global relationship between DNA copy number and gene expression in cancer cells. In agreement with our findings, Phillips *et al.* (14) have shown that with the acquisition of tumorigenicity in an immortalized prostate epithelial cell line, new chromosomal gains and losses resulted in a statistically significant respective increase and decrease in the average expression level of involved genes. In contrast, Platzer *et al.* (15) recently reported that in metastatic colon tumors only ~4% of genes within amplified regions were found more highly (>2-fold) expressed, when compared with normal colonic epithelium. This report differs substantially from our finding that 62% of highly amplified genes in breast cancer exhibit at least 2-fold increased expression. These contrasting findings may reflect methodological differences between the

studies. For example, the study of Platzer *et al.* (15) may have systematically under-measured gene expression changes. In this regard it is remarkable that only 14 transcripts of many thousand residing within unamplified chromosomal regions were found to exhibit at least 4-fold altered expression in metastatic colon cancer. Additionally, their reliance on lower-resolution chromosomal CGH may have resulted in poorly delimiting the boundaries of high-complexity amplicons, effectively overcalling regions with amplification. Alternatively, the contrasting findings for amplified genes may represent real biological differences between breast and metastatic colon tumors; resolution of this issue will require further studies.

Our finding that widespread DNA copy number alteration has a large, pervasive and direct effect on global gene expression patterns in breast cancer has several important implications. First, this finding supports a high degree of copy number-dependent gene expression in tumors. Second, it suggests that most genes are not subject to specific autoregulation or dosage compensation. Third, this finding cautions that elevated expression of an amplified gene cannot alone be considered strong independent evidence of a candidate oncogene's role in tumorigenesis. In our study, fully 62% of highly amplified genes demonstrated moderately or highly elevated expression. This highlights the importance of high-resolution mapping of amplicon boundaries and shape [to identify the "driving" gene(s) within amplicons (16)], on a large number of samples, in addition to functional studies. Fourth, this finding suggests that analyzing the genomic distribution of expressed genes, even within existing microarray gene expression data sets, may permit the inference of DNA copy number aberration, particularly aneuploidy (where gene expression can be averaged across large chromosomal regions; see Fig. 3 and supporting information). Fifth, this finding implies that a substantial portion of the phenotypic uniqueness (and by extension, the heterogeneity in clinical behavior) among patients' tumors may be traceable to underlying variation in DNA copy number. Sixth, this finding supports a possible role for widespread DNA copy number alteration in tumorigenesis (17, 18), beyond the amplification of specific oncogenes and deletion of specific tumor suppressor genes. Widespread DNA copy number alteration, and the concomitant widespread imbalance in gene expression, might disrupt critical stochiometric relationships in cell metabolism and physiology (e.g., proteosome, mitotic spindle), possibly promoting further chromosomal instability and directly contributing to tumor development or progression. Finally, our findings suggest the possibility of cancer therapies that exploit specific or global imbalances in gene expression in cancer.

We thank the many members of the P.O.B. and D.B. labs for helpful discussions. J.R.P. was a Howard Hughes Medical Institute Physician Postdoctoral Fellow during a portion of this work. P.O.B. is a Howard Hughes Medical Institute Associate Investigator. This work was supported by grants from the National Institutes of Health, the Howard Hughes Medical Institute, the Norwegian Cancer Society, and the Norwegian Research Council.

1. Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F. & Pinkel, D. (1992) *Science* 258, 818–821.
2. Kallioniemi, A., Kallioniemi, O. P., Piper, J., Tanner, M., Stokke, T., Chen, L., Smith, H. S., Pinkel, D., Gray, J. W. & Waldman, F. M. (1994) *Proc. Natl. Acad. Sci. USA* 91, 2156–2160.
3. Tirkkonen, M., Tanner, M., Karhu, R., Kallioniemi, A., Isola, J. & Kallioniemi, O. P. (1998) *Genes Chromosomes Cancer* 21, 177–184.
4. Forozan, F., Mahlamaki, E. H., Monni, O., Chen, Y., Veldman, R., Jiang, Y., Gooden, G. C., Ethier, S. P., Kallioniemi, A. & Kallioniemi, O. P. (2000) *Cancer Res.* 60, 4519–4525.
5. Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T. & Lichter, P. (1997) *Genes Chromosomes Cancer* 20, 399–407.
6. Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., *et al.* (1998) *Nat. Genet.* 20, 207–211.
7. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D. & Brown, P. O. (1999) *Nat. Genet.* 23, 41–46.
8. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., *et al.* (2000) *Nature (London)* 406, 747–752.
9. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* 98, 10869–10874.
10. Schuler, G. D. (1997) *J. Mol. Med.* 75, 694–698.
11. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) *Nature (London)* 409, 860–921.
12. Fejzo, M. S., Godfrey, T., Chen, C., Waldman, F. & Gray, J. W. (1998) *Genes Chromosomes Cancer* 22, 105–113.
13. Hughes, T. R., Roberts, C. J., Dai, H., Jones, A. R., Meyer, M. R., Slade, D., Burchard, J., Dow, S., Ward, T. R., Kidd, M. J., *et al.* (2000) *Nat. Genet.* 25, 333–337.
14. Phillips, J. L., Hayward, S. W., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., Pezullo, J. R., Ghadimi, B. M., Grossfeld, G. D., Rivera, A., *et al.* (2001) *Cancer Res.* 61, 8143–8149.
15. Platzer, P., Upender, M. B., Wilson, K., Willis, J., Lutterbaugh, J., Nosrati, A., Willson, J. K., Mack, D., Ried, T. & Markowitz, S. (2002) *Cancer Res.* 62, 1134–1138.
16. Albertson, D. G., Ylstra, B., Segraves, R., Collins, C., Dairkee, S. H., Kowbel, D., Kuo, W. L., Gray, J. W. & Pinkel, D. (2000) *Nat. Genet.* 25, 144–146.
17. Li, R., Yerganian, G., Duesberg, P., Kraemer, A., Willer, A., Rausch, C. & Hehlmann, R. (1997) *Proc. Natl. Acad. Sci. USA* 94, 14506–14511.
18. Rasnick, D. & Duesberg, P. H. (1999) *Biochem. J.* 340, 621–630.

TECHNICAL UPDATE

FROM YOUR LABORATORY SERVICES PROVIDER

HER-2/neu Breast Cancer Predictive Testing

Julie Sanford Hanna, Ph.D. and Dan Mornin, M.D.

EACH YEAR, OVER 182,000 WOMEN in the United States are diagnosed with breast cancer, and approximately 45,000 die of the disease.[1] Incidence appears to be increasing in the United States at a rate of roughly 2% per year. The reasons for the increase are unclear, but non-genetic risk factors appear to play a large role.[2]

Five-year survival rates range from approximately 65%-85%, depending on demographic group, with a significant percentage of women experiencing recurrence of their cancer within 10 years of diagnosis. One of the factors most predictive for recurrence once a diagnosis of breast cancer has been made is the number of axillary lymph nodes to which tumor has metastasized. Most node-positive women are given adjuvant therapy, which increases their survival. However, 20%-30% of patients without axillary node involvement also develop recurrent disease, and the difficulty lies in how to identify this high-risk subset of patients. These patients could benefit from increased surveillance, early intervention, and treatment.

Prognostic markers currently used in breast cancer recurrence prediction include tumor size, histological grade, steroid hormone receptor status, DNA ploidy, proliferative index, and cathepsin D status. Expression of growth factor receptors and over-expression of the HER-2/neu oncogene have also been identified as having value regarding treatment regimen and prognosis.

HER-2/neu (also known as c-erbB2) is an oncogene that encodes a transmembrane glycoprotein that is homologous to, but distinct from, the epidermal growth factor receptor. Numerous studies have indicated that high levels of expression of this protein are associated with rapid tumor growth, certain forms of therapy resistance, and shorter disease-free survival. The gene has been shown to be amplified and/or overexpressed in 10%-30% of invasive breast cancers and in 40%-60% of intraductal breast carcinoma.[3]

There are two distinct FDA-approved methods by which HER-2/neu status can be evaluated: immunohistochemistry (IHC, HercepTest™) and FISH (fluorescent in situ hybridization, PathVysion™ Kit). Both methods can be performed on archived and current specimens. The first method allows visual assessment of the amount of HER-2/neu protein present on the cell membrane. The latter method allows direct quantification of the level of gene amplification present in the tumor, enabling differentiation between low- versus high-amplification. At least one study has demonstrated a difference in recurrence risk in women younger than 40 years of age for low- versus high-amplified tumors (54.5% compared to 85.7%); this is compared to a recurrence rate of 16.7% for patients with no HER-2/neu gene amplification.[4] HER-2/neu status may be particularly important to establish in women with small ($\leq$1 cm) tumor size.

The choice of methodology for determination of HER-2/neu status depends in part on the clinical setting. FDA approval for the Vysis FISH test was granted based on clinical trials involving 1549 node-positive patients. Patients received one of three different treatments consisting of different doses of cyclophosphamide, Adriamycin, and 5-fluorouracil (CAF). The study showed that patients with amplified HER-2/neu benefited from treatment with higher doses of adriamycin-based therapy, while those with normal HER-2/neu levels did not. The study therefore identified a sub-set of women, who because they did not benefit from more aggressive treatment, did not need to be exposed to the associated side effects. In addition, other evidence indicates that HER-2/neu amplification in node-negative patients can be used as an independent prognostic indicator for early recurrence, recurrent disease at any time and disease-related death.[5] Demonstration of HER-2/neu gene amplification by FISH has also been shown to be of value in predicting response to chemotherapy in stage-2 breast cancer patients.

Selection of patients for Herceptin© (Trastuzumab) monoclonal antibody therapy, however, is based upon demonstration of HER-2/neu protein overexpression using HercepTest™. Studies using Herceptin© in patients with metastatic breast cancer show an increase in time to disease progression, increased response rate to chemotherapeutic agents and a small increase in overall survival rate. The FISH assays have not yet been approved for this purpose, and studies looking at response to Herceptin© in patients with or without gene amplification status determined by FISH are in progress.

In general, FISH and IHC results correlate well. However, subsets of tumors are found which show discordant results; i.e., protein overexpression without gene amplification or lack of protein overexpression with gene amplification. The clinical significance of such results is unclear. Based on the above considerations, HER-2/neu testing at SHMC/PAML will utilize immunohistochemistry (HercepTest©) as a screen, followed by FISH in IHC-negative cases. Alternatively, either method may be ordered individually depending on the clinical setting or clinician preference.

CPT code information

HER-2/neu via IHC

88342 (including interpretive report)

HER-2/neu via FISH

88271×2 Molecular cytogenetics, DNA probe, each
88274 Molecular cytogenetics, interphase in situ hybridization, analyze 25-99 cells
88291 Cytogenetics and molecular cytogenetics, interpretation and report

Procedural Information

Immunohistochemistry is performed using the FDA-approved DAKO antibody kit, Herceptest©. The DAKO kit contains reagents required to complete a two-step immunohistochemical staining procedure for routinely processed, paraffin-embedded specimens. Following incubation with the primary rabbit antibody to human HER-2/neu protein, the kit employs a ready-to-use dextran-based visualization reagent. This reagent consists of both secondary goat anti-rabbit antibody molecules with horseradish peroxidase molecules linked to a common dextran polymer backbone, thus eliminating the need for sequential application of link antibody and peroxidase conjugated antibody. Enzymatic conversion of the subsequently added chromogen results in formation of visible reaction product at the antigen site. The specimen is then counterstained; a pathologist using light-microscopy interprets results.

FISH analysis at SHMC/PAML is performed using the FDA-approved PathVysion™ HER-2/neu DNA probe kit, produced by Vysis, Inc. Formalin fixed, paraffin-embedded breast tissue is processed using routine histological methods, and then slides are treated to allow hybridization of DNA probes to the nuclei present in the tissue section. The Pathvysion™ kit contains two direct-labeled DNA probes, one specific for the alphoid repetitive DNA (CEP 17, spectrum orange) present at the chromosome 17 centromere and the second for the HER-2/neu oncogene located at 17q11.2-12 (spectrum green). Enumeration of the probes allows a ratio of the number of copies of chromosome 17 to the number of copies of HER-2/neu to be obtained; this enables quantification of low versus high amplification levels, and allows an estimate of the percentage of cells with HER-2/neu gene amplification. The clinically relevant distinction is whether the gene amplification is due to increased gene copy number on the two chromosome 17 homologues normally present or an increase in the number of chromosome 17s in the cells. In the majority of cases, ratio equivalents less than 2.0 are indicative of a normal/negative result, ratios of 2.1 and over indicate that amplification is present and to what degree. Interpretation of this data will be performed and reported from the Vysis-certified Cytogenetics laboratory at SHMC.

References

1. Wingo, P.A., Tong, T., Bolden, S., "Cancer Statistics", 1995;45:1:8-31.
2. "Cancer Rates and Risks", 4th ed., National Institutes of Health, National Cancer Institute, 1996, p. 120.
3. Slamon, D.J., Clark, G.M., Song, S.G., Levin, W.J., Ullrich, A., McGuire, W.L. "Human breast Cancer: Correlation of relapse and survival with amplification of the her-2/neu oncogene". Science, 235:177-182, 1987.
4. Xing, W.R., Gilchrist, K.W., Harris, C.P., Samson, W., Meisner, L.F. "FISH detection of HER-s/neu oncogene amplification in early onset breast cancer". Breast Cancer Res. And Treatment 39(2):203-212, 1996.
5. Press, M.F. Bernstein, L., Thomas, P.A., Meisner, L.F., Zhou, J.Y., Ma, Y., Hung, G., Robinson, R.A., Harris, C., El-Naggar, A., Slamon, D.J., Phillips, R.N., Ross, J.S., Wolman, S.R., Flom, K.J., "Her-2/neu gene amplification characterized by fluorescence in situ hybridization: poor prognosis in node-negative breast carcinomas", J. Clinical Oncology 15(8):2894-2904, 1997.

Provided for the clients of

PATHOLOGY ASSOCIATES MEDICAL LABORATORIES
PACLAB NETWORK LABORATORIES
TRI-CITIES LABORATORY
TREASURE VALLEY LABORATORY

For more information, please contact your local representative.

Review

Paul A. Haynes
Steven P. Gygi
Daniel Figeys
Ruedi Aebersold

Department of Molecular Biotechnology, University of Washington, Seattle, WA, USA

Proteome analysis: Biological assay or data archive?

In this review we examine the current state of proteome analysis. There are three main issues discussed: why it is necessary to study proteomes; how proteomes can be analyzed with current technology; and how proteome analysis can be used to enhance biological research. We conclude that proteome analysis is an essential tool in the understanding of regulated biological systems. Current technology, while still mostly limited to the more abundant proteins, enables the use of proteome analysis both to establish databases of proteins present, and to perform biological assays involving measurement of multiple variables. We believe that the utility of proteome analysis in future biological research will continue to be enhanced by further improvements in analytical technology.

Contents

1 Introduction

A proteome has been defined as the protein complement expressed by the genome of an organism, or, in multicellular organisms, as the protein complement expressed by a tissue or differentiated cell [1]. In the most common implementation of proteome analysis the proteins extracted from the cell or tissue analyzed are separated by high resolution two-dimensional gel electrophoresis (2-DE), detected in the gel and identified by their amino acid sequence. The ease, sensitivity and speed with which gel-separated proteins can be identified by the use of recently developed mass spectrometric techniques have dramatically increased the interest in proteome technology. One of the most attractive features of such analyses is that complex biological systems can potentially be studied in their entirety, rather than as a multitude of individual components. This makes it far easier to uncover the many complex, and often obscure, relationships between mature gene products in cells. Large-scale proteome characterization projects have been undertaken for a number of different organisms and cell types. Microbial proteome projects currently in progress include, for example: *Saccharomyces cerevisiae* [2], *Salmonella enterica* [3], *Spiroplasma melliferum* [4], *Mycobacterium tuberculosis* [5], *Ochrobactrum anthropi* [6], *Haemophilus influenzae* [7], *Synechocystis spp.* [8], *Escherichia coli* [9], *Rhizobium leguminosarum* [10], and *Dictyostelium discoideum* [11]. Proteome projects underway for tissues of more complex organisms include those for: human bladder squamous cell carcinomas [12], human liver [13], human plasma [13], human keratinocytes [12], human fibroblasts [12], mouse kidney [12], and rat serum [14]. In this manuscript we critically assess the concept of proteome analysis and the technical feasibility of establishing complete proteome maps, and discuss ways in which proteome analysis and biological research intersect.

Correspondence: Professor Ruedi Aebersold, Department of Molecular Biotechnology, University of Washington, Box 357730, Seattle, WA, 98195, USA (Tel: +206-685-4235; Fax: +206-685-6392; E-mail: ruedi@u.washington.edu)

Abbreviations: CID, collision-induced dissociation; MS/MS, tandem mass spectrometry; SAGE, serial analysis of gene expression

Keywords: Proteome / Two-dimensional polyacrylamide gel electrophoresis / Tandem mass spectrometry

2 Rationale for proteome analysis

The dramatic growth in both the number of genome projects and the speed with which genome sequences are being determined has generated huge amounts of sequence information, for some species even complete genomic sequences ([15–17]). The description of the state of a biological system by the quantitative measurement of system components has long been a primary objective in molecular biology. With recent technical advances including the development of differential display-PCR [18], cDNA microarray and DNA chip technology [19, 20] and serial analysis of gene expression (SAGE) [21, 22], it is now feasible to establish global and quantitative mRNA expression maps of cells and tissues, in which the sequence of all the genes is known, at a speed and sensitivity which is not matched by current

0173-0835/98/1111-1862 $17.50+.50/0

protein analysis technology. Given the long-standing paradigm in biology that DNA synthesizes RNA which synthesizes protein, and the ability to rapidly establish global, quantitative mRNA expression maps, the questions which arise are why technically complex proteome projects should be undertaken and what specific types of information could be expected from proteome projects which cannot be obtained from genomic and transcript profiling projects. We see three main reasons for proteome analysis to become an essential component in the comprehensive analysis of biological systems. (i) Protein expression levels are not predictable from the mRNA expression levels, (ii) proteins are dynamically modified and processed in ways which are not necessarily apparent from the gene sequence, and (iii) proteomes are dynamic and reflect the state of a biological system.

2.1 Correlation between mRNA and protein expression levels

Interpretations of quantitative mRNA expression profiles frequently implicitly or explicitly assume that for specific genes the transcript levels are indicative of the levels of protein expression. As part of an ongoing study in our laboratory, we have determined the correlation of expression at the mRNA and protein levels for a population of selected genes in the yeast *Saccharomyces cerevisiae* growing at mid-log phase (S. P. Gygi *et al.*, submitted for publication). mRNA expression levels were calculated from published SAGE frequency tables [22]. Protein expression levels were quantified by metabolic radiolabeling of the yeast proteins, liquid scintillation counting of the protein spots separated by high resolution 2-DE and mass spectrometric identification of the protein(s) migrating to each spot. The selected 80 samples constitute a relatively homogeneous group with respect to predicted half-life and expression level of the protein products. Thus far, we have found a general trend but no strong correlation between protein and transcript levels (Fig. 1). For some genes studied equivalent mRNA transcript levels translated into protein abundances which varied by more than 50-fold. Similarly, equivalent steady-state protein expression levels were maintained by transcript levels varying by as much as 40-fold (S. P. Gygi *et al.*, submitted). These results suggests that even for a population of genes predicted to be relatively homogeneous with respect to protein half-life and gene expression, the protein levels cannot be accurately predicted from the level of the corresponding mRNA transcript.

2.2 Proteins are dynamically modified and processed

In the mature, biologically active form many proteins are post-translationally modified by glycosylation, phosphorylation, prenylation, acylation, ubiquitination or one or more of many other modifications [23] and many proteins are only functional if specifically associated or complexed with other molecules, including DNA, RNA, proteins and organic and inorganic cofactors. Frequently, modifications are dynamic and reversible and may alter the precise three-dimensional structure and the state of activity of a protein. Collectively, the state of modification of the proteins which constitute a biological system

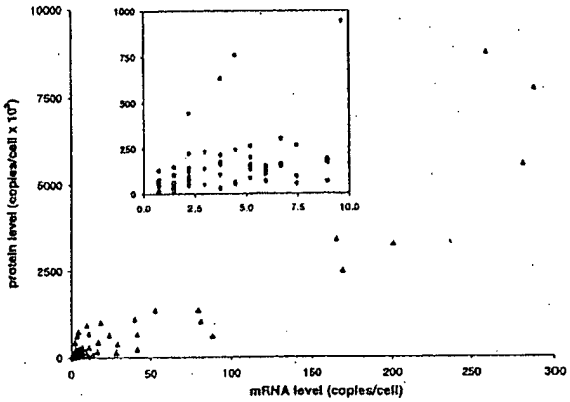


Figure 1. Correlation between mRNA and protein levels in yeast cells. For a selected population of 80 genes, protein levels were measured by ^{35}-S-radiolabeling and mRNA levels were calculated from published SAGE tables. Inset: expanded view of the low abundance region. For more experimental details, also see Figs. 5 and 6, (S. P. Gygi *et al.*, submitted).

are important indicators for the state of the system. The type of protein modification and the sites modified at a specific cellular state can usually not be determined from the gene sequence alone.

2.3 Proteomes are dynamic and reflect the state of a biological system

A single genome can give rise to many qualitatively and quantitatively different proteomes. Specific stages of the cell cycle and states of differentiation, responses to growth and nutrient conditions, temperature and stress, and pathological conditions represent cellular states which are characterized by significantly different proteomes. The proteome, in principle, also reflects events that are under translational and post-translational control. It is therefore expected that proteomics will be able to provide the most precise and detailed molecular description of the state of a cell or tissue, provided that the external conditions defining the state are carefully determined. In answer to the question of whether the study of proteomes is necessary for the analysis of biomolecular systems, it is evident that the analysis of mature protein products in cells is essential as there are numerous levels of control of protein synthesis; degradation, processing and modification, which are only apparent by direct protein analysis.

3 Description and assessment of current proteome analysis technology

3.1 Technical requirements of proteome technology

In biological systems the level of expression as well as the states of modification, processing and macro-molecular association of proteins are controlled and modulated depending on the state of the system. Comprehensive analysis of the identity, quantity and state of modification of proteins therefore requires the detection and

quantitation of the proteins which constitute the system, and analysis of differentially processed forms. There are a number of inherent difficulties in protein analysis which complicate these tasks. First, proteins cannot be amplified. It is possible to produce large amounts of a particular protein by over-expression in specific cell systems. However, since many proteins are dynamically post-translationally modified, they cannot be easily amplified in the form in which they finally function in the biological system. It is frequently difficult to purify from the native source sufficient amounts of a protein for analysis. From a technological point of view this translates into the need for high sensitivity analytical techniques. Second, many proteins are modified and processed post-translationally. Therefore, in addition to the protein identity, the structural basis for differentially modified isoforms also needs to be determined. The distribution of a constant amount of protein over several differentially modified isoforms further reduces the amount of each species available for analysis. The complexity and dynamics of post-translational protein editing thus significantly complicates proteome studies. Third, proteins vary dramatically with respect to their solubility in commonly used solvents. There are few, if any, solvent conditions in which all proteins are soluble and which are also compatible with protein analysis. This makes the development of protein purification methods particularly difficult since both protein purification and solubility have to be achieved under the same conditions. Detergents, in particular sodium dodecyl sulfate (SDS), are frequently added to aqueous solvents to maintain protein solubility. The compatibility with SDS is a big advantage of SDS polyacrylamide gel electrophoresis (SDS-PAGE) over other protein separation techniques. Thus, SDS-PAGE and two-dimensional gel electrophoresis, which also uses SDS and other detergents, are the most general and preferred methods for the purification of small amounts of proteins, provided that activity does not necessarily need to be maintained. Lastly, the number of proteins in a given cell system is typically in the thousands. Any attempt to identify and categorize all of these must use methods which are as rapid as possible to allow completion of the project within a reasonable time frame. Therefore, a successful, general proteomics technology requires high sensitivity, high throughput, the ability to differentiate differentially modified proteins, and the ability to quantitatively display and analyze all the proteins present in a sample.

3.2 2-D electrophoresis – mass spectrometry: a common implementation of proteome analysis

The most common currently used implementation of proteome analysis technology is based on the separation of proteins by two-dimensional (IEF/SDS-PAGE) gel electrophoresis and their subsequent identification and analysis by mass spectrometry (MS) or tandem mass spectrometry (MS/MS). In 2-DE, proteins are first separated by isoelectric focusing (IEF) and then by SDS-PAGE, in the second, perpendicular dimension. Separated proteins are visualized at high sensitivity by staining or autoradiography, producing two-dimensional arrays of proteins. 2-DE gels are, at present, the most commonly used means of global display of proteins in complex samples. The separation of thousands of proteins has been achieved in a single gel [24, 25] and differentially modified proteins are frequently separated. Due to the compatibility of 2-DE with high concentrations of detergents, protein denaturants and other additives promoting protein solubility, the technique is widely used.

The second step of this type of proteome analysis is the identification and analysis of separated proteins. Individual proteins from polyacrylamide gels have traditionally been identified using *N*-terminal sequencing [26, 27], internal peptide sequencing [28, 29], immunoblotting or comigration with known proteins [30]. The recent dramatic growth of large-scale genomic and expressed sequence tag (EST) sequence databases has resulted in a fundamental change in the way proteins are identified by their amino acid sequence. Rather than by the traditional methods described above, protein sequences are now frequently determined by correlating mass spectral or tandem mass spectral data of peptides derived from proteins, with the information contained in sequence databases [31–33].

There are a number of alternative approaches to proteome analysis currently under development. There is considerable interest in developing a proteome analysis stragegy which bypasses 2-DE altogether, because it is considered a relatively slow and tedious process, and because of perceived difficulties in extracting proteins from the gel matrix for analysis. However, 2-DE as a starting point for proteome analysis has many advantages compared to other techniques available today. The most significant strengths of the 2-DE-MS approach include the relatively uniform behavior of proteins in gels, the ability to quantify spots and the high resolution and simultaneous display of hundreds to thousands of proteins within a reasonable time frame.

A schematic diagram of a typical procedure of the identification of gel-separated proteins is shown in Fig. 2. Protein spots detected in the gel are enzymatically or chemically fragmented and the peptide fragments are isolated for analysis, as already indicated, most frequently by MS or MS/MS. There are numerous protocols for the generation of peptide fragments from gel-separated proteins. They can be grouped into two categories, digestion in the gel slice [28, 34] or digestion after electrotransfer out of the gel onto a suitable membrane ([29, 35–37] and reviewed in [38]). In most instances either technique is applicable and yields good results. The analysis of MS or MS/MS data is an important step in the whole process because MS instruments can generate an enormous amount of information which cannot easily be managed manually. Recently, a number of groups have developed software systems dedicated to the use of peptide MS and MS/MS spectra for the identification of proteins. Proteins are identified by correlating the information contained in the MS spectra of protein digests or MS/MS spectra of individual peptides with data contained in DNA or protein sequence databases.

The systems we are currently using in our laboratory are based on the separation of the peptides contained in protein digests by narrow bore or capillary liquid chromatog-

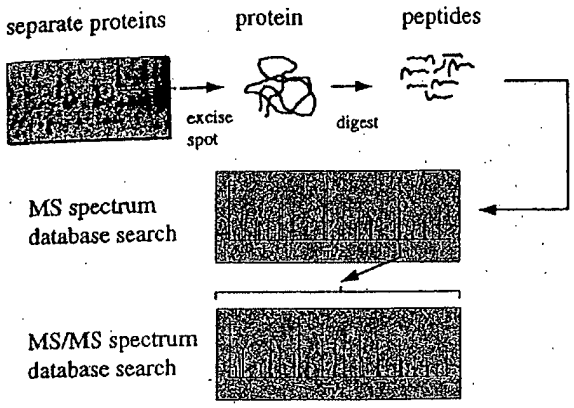


Figure 2. Schematic diagram of a procedure for identification of gel-separated proteins. Peptides can either be separated by a technique such as LC or CE, or infused as a mixture and sorted in the MS. Database searching can either be performed on peptide masses from an MS spectrum, peptide fragment masses from CID spectra of peptides, or a combination of both.

raphy [39, 40] or capillary electrophoresis [41], the analysis of the separated peptides by electrospray ionization (ESI) MS/MS, and the correlation of the generated peptide spectra with sequence databases using the SEQUEST program developed at the University of Washington [32, 33]. The system automatically performs the following operations: a particular peptide ion characterized by its mass-to-charge ratio is selected in the MS out of all the peptide ions present in the system at a particular time; the selected peptide ion is collided in a collision cell with argon (collision-induced dissociation, CID) and the masses of the resulting fragment ions are determined in the second sector of the tandem MS; this experimentally determined CID spectrum is then correlated with the CID spectra predicted from all the peptides in a sequence database which have essentially the same mass as the peptide selected for CID; this correlation matches the isolated peptide with a sequence segment in a database and thus identifies the protein from which the peptide was derived. There are a number of alternative programs which use peptide CID spectra for protein identification, but we use the SEQUEST system because it is currently the most highly automated program and has proven to be successful, versatile and robust.

3.3 Protein identification by LC-MS/MS, capillary LC-MS/MS and CE-MS/MS

It has been demonstrated repeatedly that MS has a very high intrinsic sensitivity. For the routine analysis of gel-separated proteins at high sensitivity, the most significant challenge is the handling of small amounts of sample. The crux of the problem is the extraction and transferal of peptide mixtures generated by the digestion of low nanogram amounts of protein, from gels into the MS/MS system without significant loss of sample or introduction of unwanted contaminants. We employ three different systems for introducing gel-purified samples into an MS, depending on the level of sensitivity required. As an approximate guideline, for samples containing tens of picomoles of peptides, LC-MS/MS is most appropriate; for samples containing low picomole amounts to high femtomole amounts we use capillary LC-MS/MS; and for samples containing femtomoles or less, CE-MS/MS is the method of choice.

3.3.1 LC-MS/MS

The coupling of an MS to an HPLC system using a 0.5 mm diameter or bigger reverse phase (RP) column has been described in detail [42]. This system has several advantages if a large number of samples are to be analyzed and all are available in sufficient quantity. The LC-MS and database searching program can be run in a fully automated mode using an autosampler, thus maximizing sample throughput and minimizing the need for operator interference. The relatively large column is tolerant of high levels of impurities from either gel preparation or sample matrix. Lastly, if configured with a flow-splitter and micro-sprayer [40], analyses can be performed on a small fraction of the sample (less than 5%) while the remainder of the sample is recovered in very pure solvents. This latter feature is particularly useful when an orthogonal technique is also used to analyze peptide fractions, such as scintillation of an introduced radiolabel, and this data can be correlated with peptides identified by CID spectra.

3.3.2 Capillary LC-MS

An increase of sensitivity of approximately tenfold can be achieved by using a capillary LC system with a 100 μm ID column rather than a 0.5 mm ID column as referred to above. Since very low flow rates are required for such columns, most reports have used a precolumn flow splitting system for producing solvent gradients. We have recently desribed the design and construction of a novel gradient mixing system which enables the formation of reproducible gradients at very low flow rates (low nL/min) without the need for flow splitting (A. Ducret *et al.*, submitted for publication). Using this capillary LC-MS/MS system we were able to identify gel-separated proteins if low picomole to high femtomole amounts were loaded onto the gel [40]. This system is as yet not automated and, like all capillary LC systems, is prone to blockage of the columns by microparticulates when analyzing gel-separated proteins.

3.3.3 CE-MS/MS

The highest level of sensitivity for analyzing gel-separated proteins can be achieved by using capillary electrophoresis – mass spectrometry (CE-MS). We have described in the past a solid-phase extraction capillary electrophoresis (SPE-CE) system which was used with triple quadrupole and ion trap ESI-MS/MS systems for the identification of proteins at the low femtomole to subfemtomole sensitivity level [43, 44]. While this system is highly sensitive, its operation is labor-intensive and its operation has not been automated. In order to devise an analytical system with both the sensitivity of a CE and the level of automation of LC, we have constructed

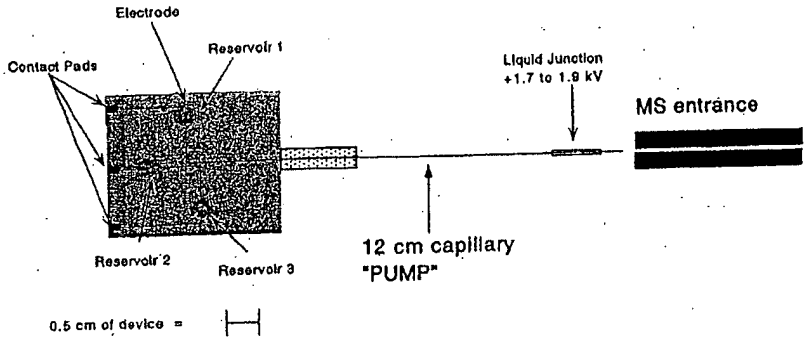


Figure 3. Schematic illustration of a microfabricated analytical system for CE, consisting of a micromachined device, coated capillary electroosmotic pump, and microelectrospray interface. The dimensions of the channels and reservoir are as indicated in the text. The channels on the device were graphically enhanced to make them more visible. Reproduced from [45], with permission.

microfabricated devices for the introduction of samples into ESI-MS for high-sensitivity peptide analysis.

The basic device is a piece of glass into which channels of 10–30 μm in depth and 50–70 μm in diameter are etched by using photolithography/etching techniques similar to the ones used in the semiconductor industry. (A simple device is shown in Fig. 3). The channels are connected to an external high voltage power supply [45]. Samples are manipulated on the device and off the device to the MS by applying different potentials to the reservoirs. This creates a solvent flow by electroosmotic pumping which can be redirected by changing the position of the electrode. Therefore, without the need for valves or gates and without any external pumping, the flow can be redirected by simply switching the position of the electrodes on the device. The direction and rate of the flow can be modulated by the size and the polarity of the electric field applied and also by the charge state of the surface.

The type of data generated by the system is illustrated in Fig. 4, which shows the mass spectrum of a peptide sample representing the tryptic digest of carbonic anhydrase at 290 fmol/μL. Each numbered peak indicates a peptide successfully identified as being derived from carbonic anhydrase. Some of the unassigned signals may be chemical or peptide contaminants. The MS is programmed to automatically select each peak and subject the peptide to CID. The resulting CID spectra are then used to identify the protein by correlation with sequence databases. Therefore, this system allows us to concurrently apply a number of protein digests onto the device, to sequentially mobilize the samples, to automatically generate CID spectra of selected peptide ions and to search sequence databases for protein identification. These steps are performed automatically without the need for user input and proteins can be identified at very low femtomole level sensitivity at a rate of approximately one protein per 15 min.

3.4 Assessment of 2-DE-MS proteome technology

Using a combination of the analytical techniques described above we have identified the 80 protein spots indicated in Fig. 5. The protein pattern was generated by separating a total of 40 microgram of protein contained in a total cell lysate of the yeast strain YPH499 by high resolution 2-DE and silver staining of the separated proteins. To estimate how far this type of proteome analysis can penetrate towards the identification of low abundance proteins, we have calculated the codon bias of the genes encoding the respective proteins. Codon bias is a

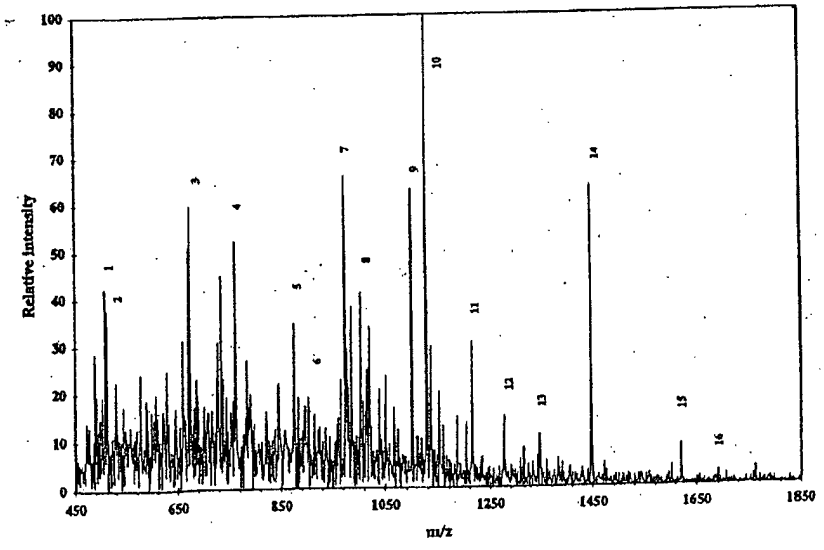


Figure 4. MS spectrum of a tryptic digest of carbonic anhydrase using the microfabricated system shown in Fig. 3. 290 fmol/μL of carbonic anhydrase tryptic digest was infused into a Finnigan LCQ ion trap MS. Each peak was selected for CID, and those which were identified as containing peptides derived from carbonic anhydrase are numbered. Reproduced from [45], with permission.

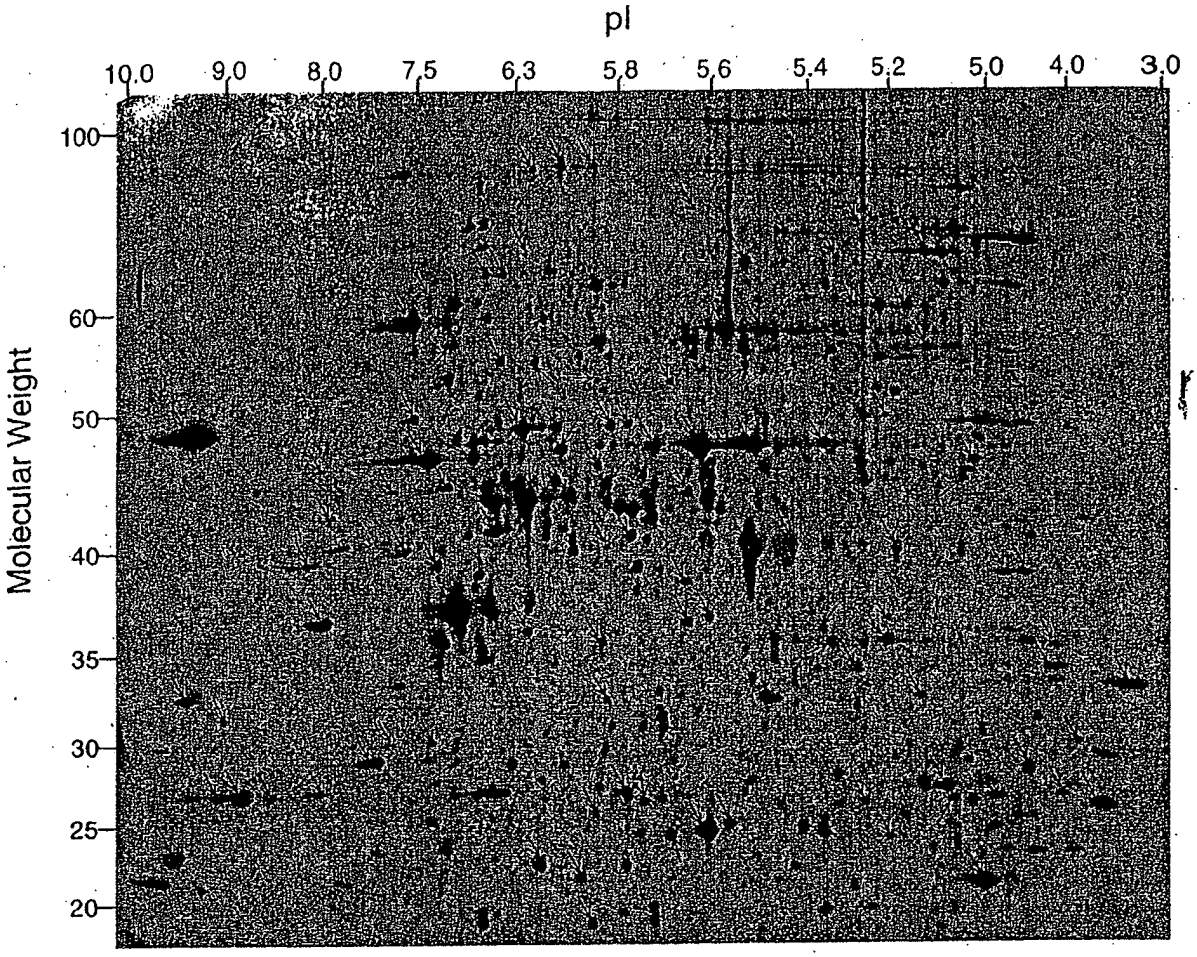


Figure 5. 2-DE separation of a lysate of yeast cells, with identified proteins highlighted. The first dimension of separation was an IPG from pH 3–10, and the second dimension was a 10%T SDS-PAGE gel. Proteins were visualized by silver staining. Further details of experimental procedures are included in S. P. Gygi *et al.* (submitted).

calculated measure of the degree of redundancy of triplet DNA codons used to produce each amino acid in a particular gene sequence. It has been shown to be a useful indicator of the level of the protein product of a particular gene sequence present in a cell [46]. The general rule which applies is that the higher the value of the codon bias calculated for a gene, the more abundant the protein product of that gene becomes. The calculated codon bias values corresponding to the proteins identified in Fig. 5 are shown in Fig. 6b. Nearly all of the proteins identified (> 95%) have codon bias values of > 0.2, indicating they are highly abundant in cells. In contrast, codon bias values calculated for the entire yeast genome (Fig. 6a) show that the majority of proteins present in the proteome have a codon bias of < 0.2 and are thus of low abundance.

This finding is of considerable importance in our assessment of the current status of proteome analysis technology. It is clear that even using highly sensitive analytical techniques, we are only able to visualize and identify the more abundant proteins. Since many important regulatory proteins are present only at low abundance, these would not be amenable to analysis using such techniques. This situation would be exacerbated in the analysis of proteomes containing many more proteins than the approximately 6000 gene products present in yeast cells [16]. In the analysis of, for example, the proteome of any human cells, there are potentially 50 000–100 000 gene products [47]. Inherent limitations on the amount of protein that can be loaded on 2-DE, and the number of components that can be resolved, indicate that only the most highly abundant fraction of the many gene products could be successfully analyzed. One approach that has been employed to circumvent these limitations is the use of very narrow range immobilized pH gradient strips for the first-dimension separation of 2-DE [48]. Since only those proteins which focus within the narrow range will enter the second dimension of separation, a much higher sample loading within the desired range is possible. This, in turn, can lead to the visualization and identification of less abundant proteins.

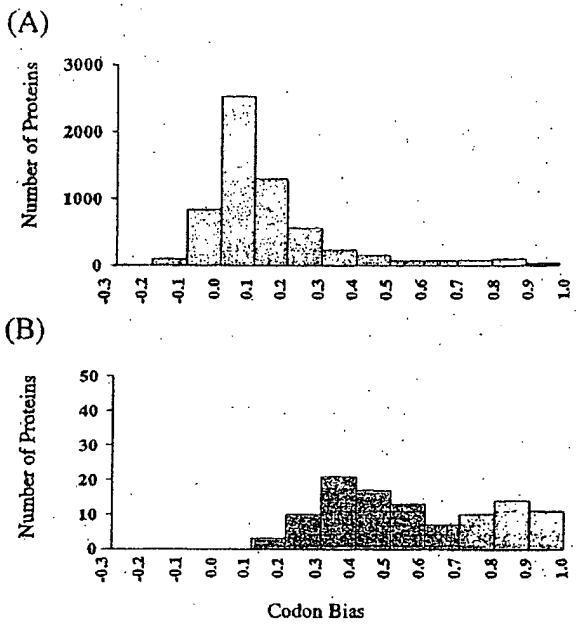


Figure 6. Calculated codon bias values for yeast proteins. (A) Distribution of calculated values for the entire yeast proteome. (B) Distribution of calculated values for the subset of 80 identified proteins also shown in Figs. 1 and 5. Further details of experimental procedures are included in S. P. Gygi *et al.* (submitted).

4 Utility of proteome analysis for biological research

For the success of proteomics as a mainstream approach to the analysis of biological systems it is essential to define how proteome analysis and biological research projects intersect. Without a clear plan for the implementation of proteome-type approaches into biological research projects the full impact of the technology can not be realized. The literature indicates that proteome analysis is used both as a database/data archive, and as a biological assay or biological research tool.

4.1 The proteome as a database

The use of proteomics as a database or data archive essentially entails an attempt to identify all the proteins in a cell or species and to annotate each protein with the known biological information that is relevant for each protein. The level of annotation can, of course, be extensive. The most common implementation of this idea is the separation of proteins by high resolution 2-DE, the identification of each detected protein spot and the annotation of the protein spots in a 2-DE gel database format. This approach is complicated by the fact that it is difficult to precisely define a proteome and to decide which proteome should be represented in the database. In contrast to the genome of a species, which is essentially static, the proteome is highly dynamic. Processes such as differentiation, cell activation and disease can all significantly change the proteome of a species. This is illustrated in Fig. 7. The figure shows two high-resolution 2-DE maps of proteins isolated from rat serum. Fig. 7A is from the serum of normal rats, while Fig. 7B is from the serum of rats in acute-phase serum after prior treatment with an inflammation-causing agent [49]. It is obvious that the protein patterns are significantly different in several areas, raising the question of exactly which proteome is being described.

Therefore, a comprehensive proteome database of a species or cell type needs to contain all of the parameters which describe the state and the type of the cells from which the proteins were extracted as well as the software tools to search the database with queries which reflect the dynamics of biological systems. A comprehensive proteome database should be capable of quantitatively describing the fate of each protein if specific systems and pathways are activated in the cell. Specifically, the quantity, the degree of modification, the subcellular location and the nature of molecules specifically interacting with a protein as well as the rate of change of these variables should be described. Using these admittedly stringent criteria, there is currently no comlete proteome database. A number of such databases are, however, in the process of being constructed. The most advanced among them, in our opinion, are the yeast protein database YPD [50] (accessible at http://www.ypd.com) and the human 2D-PAGE databases of the Danish Centre for Human Genome Research [12] (accessible at http://biobase.dk/cgi-bin/celis). While neither can be considered complete as not all of the potential gene products are identified, both contain extensive annotation of supplemental information for many of the spots which are positively identified in reference samples.

4.2 The proteome as a biological assay

The use of proteome analysis as a biological assay or research tool represents an alternative approach to integrating biology with proteomics. To investigate the state of a system, samples are subjected to a specific proceess that allows the quantitative or qualitative measurement of some of the variables which describe the system. In typical biochemical assays one variable (*e.g.*, enzyme activity) of a single component (*e.g.*, a particular enzyme) is measured. Using proteomics as an assay, multiple variables (*e.g.*, expression level, rate of synthesis, phosphorylation state, etc.) are measured concurrently on many (ideally all) of the proteins in a sample. The use of proteomics as an assay is a less far-reaching proposition than the construction of a comprehensive proteome database. It does, however, represent a pragmatic approach which can be adapted to investigate specific systems and pathways, as long as the interpretation of the results takes into account that with current technology not all of the variables which describe the system can be observed (see Section 3.4).

A common implementation of proteome analysis as a biological assay is when a 2-DE protein pattern generated from the analysis of an experimental sample is compared to an array of reference patterns representing different states of the system under investigation. The state of the experimental system at the time the sample was generated is therefore determined by the quantita-

tive comparative analysis of hundreds to a few thousand proteins. Comparative analysis of the 2-DE patterns furthermore highlights quantitative and qualitative differences in the protein profiles which correlate with the state of the system. For this type of analysis it is not essential that all the proteins are identified or even visualized, although the results become more informative as more proteins are compared. It is obvious, however, that the possibility to identify any protein deemed characteristic for a particular state dramatically enhances this approach by opening up new avenues for experimentation.

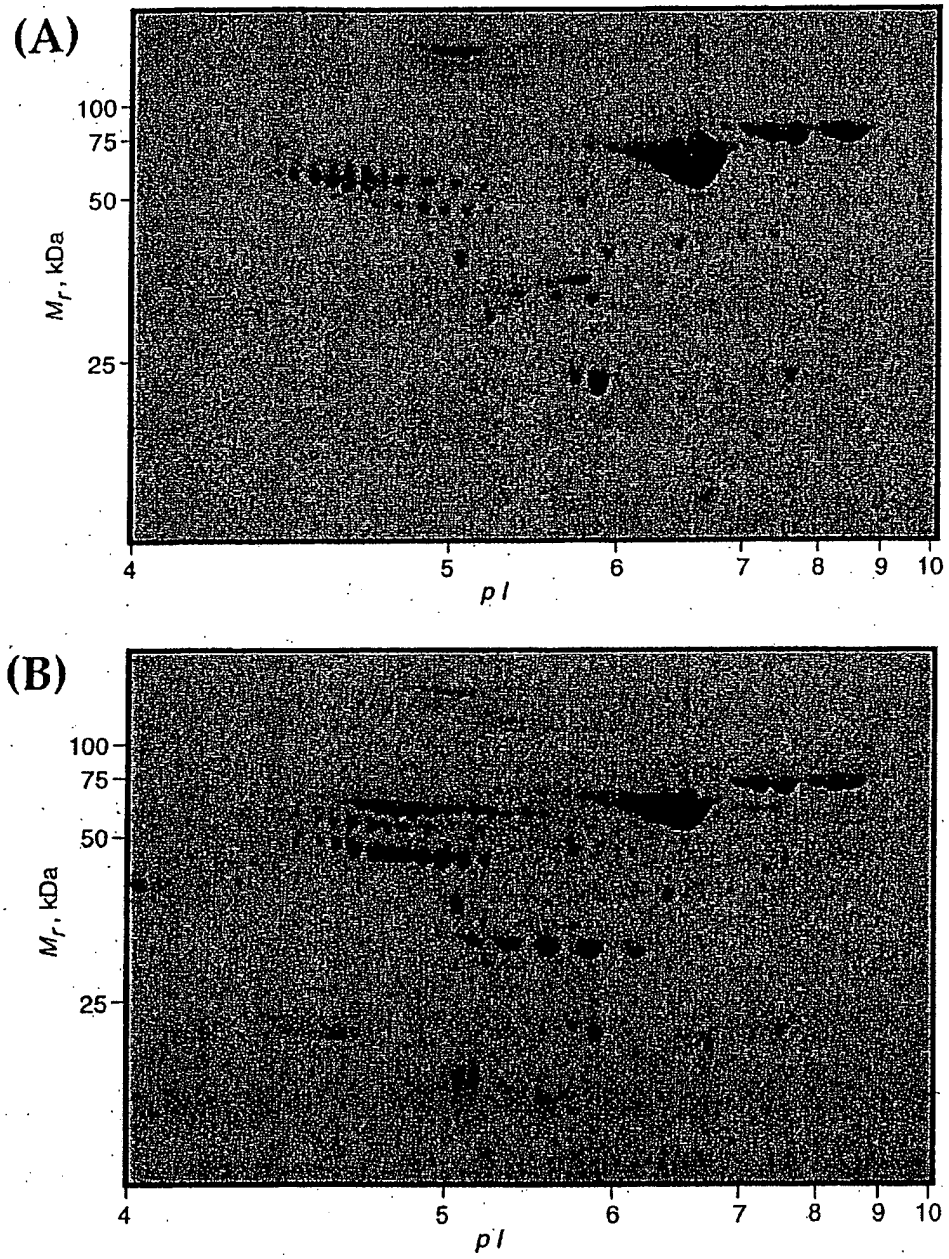


Figure 7. High resolution 2-DE map of proteins isolated from rat serum with or without prior exposure to an inflammation-causing agent. (A) normal rat serum, (B) acute-phase serum from rats which had previously been exposed to an inflammation-causing agent. The first dimension of separation is an IPG from pH 4–10, and the second dimension is a 7.5–17.5%T gradient SDS-PAGE gel. Proteins were visualized by staining with amido black. Further details of experimental procedures are included in [14, 49].

Proteome analysis as a biological assay has been successfully used in the field of toxicology, to characterize disease states or to study differential activation of cells. The approach is limited, of course, by the fact that only the visible protein spots are included in the assay, and it is well known that a substantial but far from complete fraction of cellular proteins are detected if a total cell lysate is separated by 2-DE. Proteins may not be detected in 2-DE gels because they are not abundant enough to be visualized by the detection method used, because they do not migrate within the boundaries (size, p*I*) resolved by the gel, because they are not soluble under the conditions used, or for other reasons.

A different way to use proteome analysis as a biological assay to define the state of a biological system is to take advantage of the wealth of information contained in 2-DE protein patterns. 2-DE is referred to as two-dimensional because of the electrophoretic mobility and the isoelectric points which define the position of each protein in a 2-DE pattern. In addition to the two dimensions used to generate the protein patterns, a number of additional data dimensions are contained in the protein patterns. Some of these dimensions such as protein expression level, phosphorylation state, subcellular location, association with other proteins, rate of synthesis or degradation indicate the activity state of a protein or a biological system. Comparative analysis of 2-DE protein patterns representing different states is therefore ideally suited for the detection, identification and analysis of suitable markers. Once again it must be emphasized that in this type of experiment only a fraction of the cellular proteins is analyzed. Since many regulatory proteins are of low abundance, this limitation is a concern, particularly in cases in which regulatory pathways are being investigated.

5 Concluding remarks

In this report we have addressed three main issues related to proteome analysis. First, we have discussed the rationale for studying proteomes. Second, we have assessed the technical feasibility of analyzing proteomes and described current proteome technology, and third, we have analyzed the utility of proteome analysis for biological research. It is apparent that proteome analysis is an essential tool in the analysis of biological systems. The multi-level control of protein synthesis and degradation in cells means that only the direct analysis of mature protein products can reveal their correct identities, their relevant state of modification and/or association and their amounts. Recently developed methods have enabled the identification of proteins at ever-increasing sensitivity levels and at a high level of automation of the analytical processes. A number of technical challenges, however, remain. While it is currently possible to identify essentially any protein spots that can be visualized by common staining methods, it is apparent that without prior enrichment only a relatively small and highly selected population of long-lived, highly expressed proteins is observed. There are many more proteins in a given cell which are not visualized by such methods. Frequently it is the low abundance proteins that execute key regulatory functions.

We have outlined the two principal ways proteome analysis is currently being used to intersect with biological research projects: the proteome as a database or data archive and proteome analysis as a biological assay. Both approaches have in common that at present they are conceptually and technically limited. Current proteome databases typically are limited to one cell type and one state of a cell and therefore do not account for the dynamics of biological systems. The use of proteome analysis as a biological assay can provide a wealth of information, but it is limited to the proteins detected and is therefore not truly proteome-wide. These limitations in proteomics are to a large extent a reflection of the fact that proteins in their fully processed form cannot easily be amplified and are therefore difficult to isolate in amounts sufficient for analysis or experimentation. The fact that to date no complete proteome has been described further attests to these difficulties. With continued rapid progress in protein analysis technology, however, we anticipate that the goal of complete proteome analysis will eventually become attainable.

We would like to acknowledge the funding for our work from the National Science Foundation Science and Technology Center for Molecular Biotechnology and from the NIH. We thank Yvan Rochon and Bob Franza for providing the yeast gel shown and Elisabetta Gianazza for providing the rat serum gels shown.

Received April 21, 1998

6 References

[1] Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J.-C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., Hochstrasser, D. F., *Bio/Technology* 1996, *14*, 61–65.

[2] Hodges, P. E., Payne, W. E., Garrels, J. I., *Nucleic Acids Res.* 1998, *26*, 68–72.

[3] O'Connor, C. D., Farris, M., Fowler, R., Qi, S. Y., *Electrophoresis* 1997, *18*, 1483–1490.

[4] Cordwell, S. J., Basseal, D. J., Humphery-Smith, I., *Electrophoresis* 1997, *18*, 1335–1346.

[5] Urquhart, B. L., Atsalos, T. E., Roach, D., Basseal, D. J., Bjellqvist, B., Britton, W. L., Humphery-Smith, I., *Electrophoresis* 1997, *18*, 1384–1392.

[6] Wasinger, V. C., Bjellqvist, B., Humphery-Smith, I., *Electrophoresis* 1997, *18*, 1373–1383.

[7] Link, A. J., Hays, L. G., Carmack, E. B., Yates III, J. R., *Electrophoresis* 1997, *18*, 1314–1334.

[8] Sazuka, T., Ohara, O., *Electrophoresis* 1997, *18*, 1252–1258.

[9] VanBogelen, R. A., Abshire, K. Z., Moldover, B., Olson, E. R., Neidhardt, F. C., *Electrophoresis* 1997, *18*, 1243–1251.

[10] Guerreiro, N., Redmond, J. W., Rolfe, B. G., Djordjevic, M. A., *Mol. Plant Microbe Interact.* 1997, *10*, 506–516.

[11] Yan, J. X., Tonella, L., Sanchez, J.-C., Wilkins, M. R., Packer, N. H., Gooley, A. A., Hochstrasser, D. F., Williams, K. L., *Electrophoresis* 1997, *18*, 491–497.

[12] Celis, J., Gromov, P., Ostergaard M., Madsen, P., Honoré, B., Dejgaard, K., Olsen, E., Vorum, H., Kristensen, D. B., Gromova, I., Haunso, A., Van Damme, J., Puype, M., Vandekerckhove, J., Rasmussen, H. H., *FEBS Lett.* 1996, *398*, 129–134.

[13] Appel, R. D., Sanchez, J.-C., Bairoch, A., Golaz, O., Miu, M., Vargas, J. R., Hochstrasser, D. F., *Electrophoresis* 1993, *14*, 1232–1238.

[14] Haynes, P., Miller, I., Aebersold, R., Gemeiner, M., Eberini, I., Lovati, R. M., Manzoni, C., Vignati, M., Gianazza, E., *Electrophoresis* 1998, *19*, 1484–1492.

[15] Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, C. O., Venter, J. C., *Science* 1995, *269*, 496–512.

[16] Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S. G., *Science* 1996, *274*, 546.

[17] Fraser, C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R., Lathigra, R., White, O., Ketchum, K. A., Dodson, R., Hickey, E. K., Gwinn, M., Dougherty, B., Tomb, J. F., Fleischmann, R. D., Richardson, D., Peterson, J., Kerlavage, A. R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M. D., Gocayne, J., Weidman, J., Utterback, T., Watthey, T., McDonald, L., Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M. D., Horst, K., Roberts, K., Hatch, B., Smith, H. O., Venter, J. C., *Nature* 1997, *390*, 580–586.

[18] Liang, P., Pardee, A. B., *Science* 1992, *257*, 967–971.

[19] Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., Brown, P. O., Davis, R. W., *Proc. Natl. Acad. Sci. USA* 1997, *94*, 13057–13062.

[20] Shalon, D., Smith, S. J., Brown, P. O., *Genome Res.* 1996, *6*, 639–645.

[21] Velculescu, V. E., Zhang, L., Vogelstein, B., Kinzler, K. W., *Science* 1995, *270*, 484–487.

[22] Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, D. E., Hieter, P., Vogelstein, B., Kinzler, K. W., *Cell* 1997, *88*, 243–251.

[23] Krishna, R. G., Wold, F., *Adv. Enzymol.* 1993, *67*, 265–298.

[24] Görg, A., Postel, W., Gunther, S., *Electrophoresis* 1988, *9*, 531–546.

[25] Klose, J., Kobalz, U., *Electrophoresis* 1995, *16*, 1034–1059.

[26] Matsudaira, P., *J. Biol. Chem.* 1987, *262*, 10035–10038.

[27] Aebersold, R. H., Teplow, D. B., Hood, L. E., Kent, S. B., *J. Biol. Chem.* 1986, *261*, 4229–4238.

[28] Rosenfeld, J., Capdevielle, J., Guillemot, J. C., Ferrara, P., *Anal. Biochem.* 1992, *203*, 173–179.

[29] Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E., Kent, S. B., *Proc. Natl. Acad. Sci. USA* 1987, *84*, 6970–6974.

[30] Honoré, B., Leffers, H., Madsen, P., Celis, J. E., *Eur. J. Biochem.* 1993, *218*, 421–430.

[31] Mann, M., Wilm, M., *Anal. Chem.* 1994, *66*, 4390–4399.

[32] Eng, J., McCormack, A. L., Yates III, J. R., *J. Amer. Mass Spectrom.* 1994, *5*, 976–989.

[33] Yates III, J. R., Eng, J. K., McCormack, A. L., Schieltz, D., *Anal. Chem.* 1995, *67*, 1426–1436.

[34] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.

[35] Hess, D., Covey, T. C., Winz, R., Brownsey, R. W., Aebersold, R., *Protein Sci.* 1993, *2*, 1342–1351.

[36] van Oostveen, I., Ducret, A., Aebersold, R., *Anal. Biochem.* 1997, *247*, 310–318.

[37] Lui, M., Tempst, P., Erdjument-Bromage, H., *Anal. Biochem.* 1996, *241*, 156–166.

[38] Patterson, S. D., Aebersold, R. A., *Electrophoresis* 1995, *16*, 1791–1814.

[39] Ducret, A., Foyn, Brunn, C., Bures, E. J., Marhaug, G., Husby, G. R. A., *Electrophoresis* 1996, *17*, 866–876.

[40] Haynes, P. A., Fripp, N., Aebersold, R., *Electrophoresis* 1998, *19*, 939–945.

[41] Figeys, D., Van Oostveen, I., Ducret, A., Aebersold, R., *Anal. Chem.* 1996, *68*, 1822–1828.

[42] Ducret, A., Van Oostveen, I., Eng, J. K., Yates III, J. R., Aebersold, R., *Protein Sci.* 1997, *7*, 706–719.

[43] Figeys, D., Ducret, A., Yates III, J. R., Aebersold, R., *Nature Biotech.* 1996, *14*, 1579–1583.

[44] Figeys, D., Aebersold, R., *Electrophoresis* 1997, *18*, 360–368.

[45] Figeys, D., Ning, Y., Aebersold, R., *Anal. Chem.* 1997, *69*, 3153–3160.

[46] Garrels, J. I., McLaughlin, C. S., Warner, J. R., Futcher, B., Latter, G. I., Kobayashi, R., Schwender, B., Volpe, T., Anderson, D. S., Mesquita-Fuentes, R., Payne, W. E., *Electrophoresis* 1997, *18*, 1347–1360.

[47] Schuler, G. D., Boguski, M. S., Stewart, E. A., Stein, L. D., Gyapay, G., Rice, K., White, R. E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E., Bentolila, S., Birren, B. B., Butler, A., Castle, A. B., Chiannilkulchai, N., Chu, A., Clee, C., Cowles, S., Day, P. J., Dibling, T., Drouot, N., Dunham, I., Duprat, S., Edwards, C., Fan, J.-B., Fang, N., Fizames, C., Garrett, C., Green, L., Hadley, D., Harris, M., Harrison, P., Brady, S., Hicks, A., Holloway, E., Hui, L., Hussain, S., Louis-Dit-Sully, C., Ma, J., MacGilvery, A., Mader, C., Maratukulam, A., Matise, T. C., McKusick, K. B., Morissette, J., Mungall, A., Muselet, D., Nusbaum, H. C., Page, D. C., Peck, A., Perkins, S., Piercy, M., Qin, F., Quackenbush, J., Ranby, S., Reif, T., Rozen, S., Sanders, X., She, X., Silva, J., Slonim, D. K., Soderlund, C., Sun, W.-L., Tabar, P., Thangarajah, T., Vega-Czarny, N., Vollrath, D., Voyticky, S., Wilmer, T., Wu, X., Adams, M. D., Auffray, C., Walter, N. A. R., Brandon, R., Dehejia, A., Goodfellow, P. N., Houlgatte, R., Hudson, J. R., Jr., Ide, S. E., Iorio, K. R., Lee, W. Y., Seki, N., Nagase, T., Ishikawa, K., Nomura, N., Phillips, C., Polymeropoulos, M. H., Sandusky, M., Schmitt, K., Berry, R., Swanson, K., Torres, R., Venter, J. C., Sikela, J. M., Beckmann, J. S., Weissenbach, J., Myers, R. M., Cox, D. R., James, M. R., Bentley, D., *et al. Science* 1996, *274*, 540–546.

[48] Sanchez, J.-C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M. R., Hochstrasser, D. F., *Electrophoresis* 1997, *18*, 324–327.

[49] Miller, I., Haynes, P., Gemeiner, M., Aebersold, R., Manzoni, C., Lovati, M. R., Vignati, M., Eberini, I., Gianazza, E., *Electrophoresis* 1998, *19*, 1493–1500.

[50] Garrels, J. I., *Nucleic Acids Res.* 1996, *24*, 46–49.

editorial Journal of proteome research

Do We Have Enough Biomarkers?

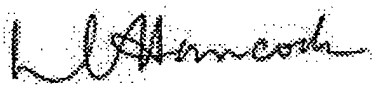
The Editor has become aware of a recent push to validate currently available biomarkers in an extensive clinical setting. The reasoning behind such a push is that there are already a significant number of biomarkers that now need to be used effectively in the clinic. Many biomarkers, such as the carcinoembryonic antigen, have been known for some time and are used widely for patient management. The older biomarkers, however, are not effective for early diagnosis.

With the advent of genomics and, later, proteomics, there has been a substantial investment in using these new tools to generate additional biomarkers. The problem with this new information is that it is too early to get consensus on what is a useful marker or what is a good patient population for such a study. Therefore, it is unclear whether the new markers currently in hand will give better clinical information than the ones that have been used in the past. An additional problem is that the markers that are generated by proteomics are not always consistent with the markers that are generated from expression profiling.

The challenge in this situation is to balance the need of patients for better, early diagnosis of disease with the need to have high-quality markers for the expensive and time-consuming validation process. This Editor believes that proteomics is at too early a stage for this new technology to have generated a quality list of markers. The risk is if we push the existing markers into extensive clinical validation, we will be missing the fruits of improvements in emerging proteomics technology. I think many people in the proteomics community would agree that federal granting agencies should be enticed to continue investments in basic proteomics technology. In addition, funding should be made available for basic science studies that will continue to generate biomarkers, and there needs to be some type of consensus-building process that can lead to a consolidation of the different lists of biomarkers.

There are good past models for such activities, such as the consensus-forming meetings that the U.S. Food and Drug Administration has held; these yielded technical innovations. One example was the generation of new protein pharmaceuticals at the advent of the biotechnology industry. Another example, in the early days of the genome sequencing program, was when a group of experts came together to agree on annotation of the early results. The Human Proteome Organization is a good example of an international group of laboratories coming together to consolidate the output from a number of studies with different technology platforms.

I would like to encourage the biomedical community not to rush to judgment in terms of biomarkers, but instead to give research more time to produce quality biomarker information. Then we should conduct a thorough evaluation of a widely agreed-on list before we attempt to determine which of these new markers are indeed worthy of extensive clinical investigation.



PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12P 21/00, C12N 1/21, 5/10, 15/00, C07H 21/04, C07K 14/46

A1

(11) International Publication Number: **WO 99/09198**

(43) International Publication Date: 25 February 1999 (25.02.99)

(21) International Application Number: PCT/US98/17211

(22) International Filing Date: 20 August 1998 (20.08.98)

(30) Priority Data:
60/056,565 21 August 1997 (21.08.97) US

(71) Applicant *(for all designated States except US)*: HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). SOPPET, Daniel, R. [US/US]; 15050 Stillfield Place, Centreville, VA 22020 (US). EBNER, Reinhard [DE/US]; 9906 Shelburne Terrace #316, Gaithersburg, MD 20878 (US).

(74) Agents: HOOVER, Kenley, K. et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: HUMAN NODAL AND LEFTY HOMOLOGUES

(57) Abstract

The present invention relates to novel Nodal and Lefty proteins which are members of the TGF–β family. In particular, isolated nucleic acid molecules are provided encoding the human Nodal and Lefty proteins. Nodal and Lefty polypeptides are also provided as are vectors, host cells and recombinant methods for producing the same. The invention further relates to screening methods for identifying agonists and antagonists of Nodal and Lefty activity. Also provided are diagnostic methods for detecting cell growth and differentiation–related disorders and therapeutic methods for treating cell growth and differentiation–related disorders.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Human Nodal and Lefty Homologues

Field of the Invention

The present invention relates to two novel human genes encoding polypeptides which are members of the transforming growth factor-beta (TGF-β) superfamily. More specifically, isolated nucleic acid molecules are provided encoding human polypeptides designated the Nodal and Lefty homologues, hereinafter referred to as "Nodal" and "Lefty", respectively. Nodal and Lefty polypeptides are also provided, as are vectors, host cells and recombinant methods for producing the same. Also provided are diagnostic methods for detecting disorders related to the regulation of cell growth and differentiation and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying agonists and antagonists of Nodal and Lefty activity.

Background of the Invention

The TGF-β family of peptide growth factors includes at least five members (TGF-β1 through TGF-β5) all of which form homodimers of approximately 25 kd. The TGF-β family belongs to a larger, extended super family of peptide signaling molecules that includes the Muellerian inhibiting substance (Cate, R. L., *et al., Cell* **45**:685-698 (1986)), decapentaplegic (Padgett, R. W., *et al., Nature* **325**:81-84 (1987)), bone morphogenic factors (Wozney, J. M., *et al., Science* **242**:1528-1534 (1988)), vg1 (Weeks, D. L. and Melton, D. A., *Cell* **51**:861-867 (1987)), activins (Vale, W., *et al., Nature* **321**:776-779 (1986)), and inhibins (Mason, A. J., *et al., Nature* **318**:659-663 (1985)). These factors are similar to TGF-β in overall structure, but share only approximately 25% amino acid identity with the TGF-β proteins and with each other. All of these molecules are thought to play an important roles in modulating growth,

development and differentiation (Kingsley, D. M. *Genes & Dev.* **8**:133-146 (1994)).

TGF-β was originally described as a factor that induced normal rat kidney fibroblasts to proliferate in soft agar in the presence of epidermal growth factor (Roberts, A. B., *et al., Proc. Natl. Acad. Sci. USA* **78**:5339-5343 (1981)). TGF-β has subsequently been shown to exert a number of different effects in a variety of cells. For example, TGF-β can inhibit the differentiation of certain cells of mesodermal origin (Florini, J. R., *et al., J. Biol. Chem.* **261**:1659-16513 (1986)), induced the differentiation of others (Seyedine, S. M. *et al., Proc. Natl. Acad. Sci. USA* **82**:2267-2271 (1985)), and potently inhibit proliferation of various types of epithelial cells, (Tucker, R. F., *Science* **226**:705-707 (1984)). This last activity has lead to the speculation that one important physiologic role for TGF-β is to maintain the repressed growth state of many types of cells. Accordingly, cells that lose the ability to respond to TGF-β are more likely to exhibit uncontrolled growth and to become tumorigenic. Indeed, cells which characteristically lack certain tumors (e.g. retinoblastoma) lack detectable TGF-β receptors at their cell surface and fail to respond to TGF-β, while their normal counterparts express self-surface receptors in their growth is potently inhibited by TGF-β (Kim Chi, A., *et al., Science* **240**:196-198 (1988)).

More specifically, TGF-β1 stimulates the anchorage-independent growth of normal rat kidney fibroblasts (Robert *et al., Proc. Natl. Acad. Sci. USA* **78**:5339-5343 (1981)). Since then it has been shown to be a multi-functional regulator of cell growth and differentiation (Sporn, *et al., Science* **233**:532-534 (1986)) being capable of such diverse effects of inhibiting the growth of several human cancer cell lines (Roberts, *et al., Proc. Natl. Acad. Sci. USA* **82**:119-123 (1985)), mouse keratinocytes, (Coffey, *et al., Cancer Res.* **48**:1596-1602 (1988)), and T and B lymphocytes (Kehrl, *et al., J. Exp. Med.* 163:1037-1050 (1986)). It also inhibits early hematopoietic progenitor cell proliferation (Goey, *et al., J.*

Immunol. **143**:877-880 (1989)), stimulates the induction of differentiation of rat muscle mesenchymal cells and subsequent production of cartilage-specific macro molecules (Seyedine, *et al., J. Biol. Chem.* **262**:1946-1949 (1986)), causes increased synthesis and secretion of collagen (Ignotz, *et al., J. Biol. Chem.* **261**:4337-4345 (1986)), stimulates bone formation (Noda, *et al., Endocrinol.* **124**:2991-2995 (1989)), and accelerates the healing of incision wounds (Mustoe, *et al., Science* **237**:1333-1335 (1987)).

Further, TGF-β1 stimulates formation of extracellular matrix molecules in the liver and lung. When levels of TGF-β1 are higher than normal, formation of fiber occurs in the extracellular matrix of the liver and lung which can be fatal. High levels of TGF-β1 occur due to chemotherapy and bone marrow transplant as an attempt to treat cancers such as breast cancer.

A second protein termed TGF-β2 was isolated from several sources including demineralized bone, a human prostatic adenocarcinoma cell line (Ikeda, *et al., J. Bio. Chem.* **26**:2406-2410 (1987)). TGF-β2 shared several functional similarities with TGF-β1. These proteins are now known to be members of a family of related growth modulatory proteins including TGF-β3 (Ten-Dijke, *et al., Proc. Natl. Acad. Sci. USA* **85**:471-4719 (1988)), Muellerian inhibitory substance and the inhibins.

Thus, there is a need for polypeptides that function as potent regulators of cell growth and differentiation, since disturbances of such regulation may be involved in disorders relating to abnormal regulation of cell growth and differentiation, cancer, tissue regeneration, and wound healing. Therefore, there is a need for identification and characterization of such human polypeptides which can play a role in detecting, preventing, ameliorating or correcting such disorders.

Summary of the Invention

The present invention provides isolated nucleic acid molecules comprising polynucleotides encoding at least a portion of the Nodal polypeptide having the complete amino acid sequence shown in SEQ ID NO:2 or the complete amino acid sequence encoded by the cDNA clone deposited as plasmid DNA as ATCC Deposit Number 209092, on June 5, 1997 or the complete amino acid sequence encoded by the cDNA clone deposited as plasmid DNA as ATCC Deposit Number 209135, on July 2, 1997. The nucleotide sequence determined by sequencing the deposited Nodal clone, which is shown in Figures 1A and B (SEQ ID NO:1), and contains a single open reading frame encoding a complete polypeptide of 283 amino acid residues initiating with a codon encoding an N-terminal aspartic acid residue at nucleotide positions 1-3 with a predicted molecular weight of about 32.5 kDa. Nucleic acid molecules of the invention include those encoding the complete amino acid sequence shown in SEQ ID NO:2, the complete amino acid sequence encoded by the cDNA clone in ATCC Deposit Numbers 209092 and 209135, which molecules also can encode additional amino acids fused to the N-terminus of the Nodal amino acid sequence.

The present invention also provides isolated nucleic acid molecules comprising polynucleotides encoding at least a portion of the Lefty polypeptide having the complete amino acid sequence shown in SEQ ID NO:4 or the complete amino acid sequence encoded by the cDNA clone deposited as plasmid DNA as ATCC Deposit Number 209091 on June 5, 1997. The nucleotide sequences determined by sequencing the deposited Lefty clone, which is shown in Figures 2A and B (SEQ ID NO:3), and contains a single open reading frame encoding a complete polypeptide of 366 amino acid residues with an initiation codon encoding an N-terminal methionine at nucleotide positions 53-55, and a predicted molecular weight of about 40.9 kDa. Nucleic acid molecules of the invention include those encoding the complete amino acid sequence shown in SEQ ID

NO:4, those encoding the complete amino acid sequence shown in SEQ ID NO:4 excluding the N-terminal methionine, the complete amino acid sequences encoded by the cDNA clone in ATCC Deposit Numbers 209091, or the complete amino acid sequences excepting the N-terminal methionine encoded by the cDNA clone in ATCC Deposit Number 209091, which molecules also can encode additional amino acids fused to the N-terminus of the Lefty amino acid sequence.

The Nodal protein of the present invention shares sequence homology with the translation product of the murine mRNA for Nodal (Figure 3; SEQ ID NO:5), including the conserved predicted active domain of about 110 amino acids. Murine Nodal is thought to be essential for mesoderm formation and subsequent organization of axial structures in early mouse development. The homology between murine Nodal and the human Nodal homologue of the present invention indicates that the human Nodal homologue of the present invention may also be involved in a developmental process such as the correct formation of various structures or in one or more post-developmental capacities including sexual development, pituitary hormone production, and the creation of bone and cartilage, as are many of the other members of the TGF-β superfamily.

The Lefty protein of the present invention shares sequence homology with the translation product of the murine mRNA for Lefty (Figure 4; SEQ ID NO:6), including the conserved predicted active domain of about 110 amino acids. Murine Lefty is thought to be important in left/right handedness of the developing organism. The homology between murine Lefty and the novel human Lefty homologue of the present invention indicates that the novel human Lefty homologue of the present invention may also be involved in correct formation of various structures with respect to the rest of the developing organism or Lefty may also be involved in one or more post-developmental capacities including sexual development, pituitary hormone production, and the creation of bone and cartilage, as are many of the other members of the TGF-β superfamily.

Nodal and Lefty polypeptides of the present invention are useful for enhancing or enriching the growth and/or differentiation of specific cell populations, e.g., embryonic cells or stem cells.

Another embodiment of the invention provides pharmaceutical compositions which contain a therapeutically effective amount of human Nodal and/or Lefty polypeptide, in a pharmaceutically acceptable vehicle or carrier. These compositions of the invention may be useful in the therapeutic modulation or diagnosis of bone, cartilage, or other connective cell or tissue growth and/or differentiation. These compositions may be used to treat such conditions as osteoarthritis, osteoporosis, and other abnormalities of bone, cartilage, muscle, tendon, ligament and/or other connective tissues and/or organs such as liver, lung, cardiac, pancreas, kidney, and other tissues. These compositions may also be useful in the growth and/or formation of cartilage, tendon, ligament, meniscus, and other connective tissues or any combination of the above (e.g., therapeutic modulation of the tendon-to-bone attachment apparatus). These compositions may also be useful in treating periodontal disease and modulating wound healing and tissue repair of such tissues as epidermis, nerve, muscle, cardiac muscle, liver, lung, cardiac, pancreas, kidney, and other tissues and/or organs. Pharmaceutical compositions containing Nodal and/or Lefty of the invention may include one or more other therapeutically useful component such as BMP-1, BMP-2, BMP-3, BMP-4, BMP-5, BMP-6, and/or BMP-7 (*See*, for example, U. S. Patent Nos. 5,108,922; 5,013,649; 5,116,738; 5,106,748; 5,187,076; and 5,141,905), BMP-8 (*See*, for example, PCT publication WO91/18098), BMP-9 (*See*, for example, PCT publication WO93/00432), BMP-10 (*See*, for example, PCT publication WO94/26893), BMP-11 (*See*, for example, PCT publication WO94/26892), BMP-12 and/or BMP-13 (*See*, for example, PCT publication WO95/16035), with other growth factors including, but not limited to, BIP, one or more of the growth and differentiation factors (GDFs), VGR-2, epidermal growth factor (EGF),

fibroblast growth factor (FGF), TGF-alpha, TGF-beta, activins, inhibins, and insulin-like growth factor (IGF).

The encoded Lefty polypeptide has a predicted leader sequence of 18 amino acids underlined in Figure 2A; and the amino acid sequence of the predicted secreted Lefty protein is also shown in Figures 2A-B, as amino acid residues 19-366 and as residues 1-348 in SEQ ID NO:4.

Thus, one embodiment of the invention provides an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding the Nodal polypeptide having the complete amino acid sequence in SEQ ID NO:2 (i.e., positions 1 to 283 of SEQ ID NO:2); (b) a nucleotide sequence encoding the predicted active Nodal polypeptide having the amino acid sequence at positions 173 to 283 of SEQ ID NO:2; (c) a nucleotide sequence encoding the Nodal polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 and/or 209135; (d) a nucleotide sequence encoding the active domain of the Nodal polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 and/or 209135; and (e) a nucleotide sequence complementary to any of the nucleotide sequences in (a), (b), (c) or (d) above.

Another embodiment of the invention provides an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding the Lefty polypeptide having the complete amino acid sequence in SEQ ID NO:4 (i.e., positions -18 to 348 of SEQ ID NO:4); (b) a nucleotide sequence encoding the Lefty polypeptide having the complete amino acid sequence in SEQ ID NO:4 excepting the N-terminal methionine (i.e., positions -17 to 348 of SEQ ID NO:4); (c) a nucleotide sequence encoding the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 60 to 348 of SEQ ID

NO:4; (d) a nucleotide sequence encoding the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 118 to 348 of SEQ ID NO:4; (e) a nucleotide sequence encoding the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 125 to 348 of SEQ ID NO:4; (f) a nucleotide sequence encoding the Lefty polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091; (g) a nucleotide sequence encoding the Lefty polypeptide having the complete amino acid sequence excepting the N-terminal methionine encoded by the cDNA clone contained in ATCC Deposit No. 209091; (h) a nucleotide sequence encoding the active domain of the Lefty polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091; and (i) a nucleotide sequence complementary to any of the nucleotide sequences in (a), (b), (c), (d), (e), (f), (g) or (h) above.

Further embodiments of the invention include isolated nucleic acid molecules that comprise a polynucleotide having a nucleotide sequence at least 90% identical, and more preferably at least 95%, 96%, 97%, 98% or 99% identical, to any of the nucleotide sequences in (a), (b), (c), (d) or (e), above, with regard to Nodal, to any of the nucleotide sequences in (a), (b), (c), (d), (e), (f), (g), (h) or (i), above, with regard to Lefty, or a polynucleotide which hybridizes, preferably under stringent hybridization conditions, to a polynucleotide in (a), (b), (c), (d) or (e), above, with regard to Nodal, or any of the nucleotide sequences in (a), (b), (c), (d), (e), (f), (g), (h) or (i), above, with regard to Lefty, listed above. This polynucleotide which hybridizes does not hybridize under stringent hybridization conditions to a polynucleotide having a nucleotide sequence consisting of only A residues or of only T residues.

An additional nucleic acid embodiment of the invention relates to an isolated nucleic acid molecule comprising a polynucleotide which encodes the

amino acid sequence of an epitope-bearing portion of a Nodal polypeptide having an amino acid sequence in (a), (b), (c), (d) or (e), with regard to Nodal, above. A further nucleic acid embodiment of the invention relates to an isolated nucleic acid molecule comprising a polynucleotide which encodes the amino acid sequence of an epitope-bearing portion of a Lefty polypeptide having an amino acid sequence in (a), (b), (c), (d), (e), (f), (g), (h) or (i), with regard to Lefty, above. A further embodiment of the invention relates to an isolated nucleic acid molecule comprising a polynucleotide which encodes the amino acid sequences of a Nodal or Lefty polypeptide having an amino acid sequence which contains at least one amino acid substitution, but not more than 50 amino acid substitutions, even more preferably, not more than 40 amino acid substitutions, still more preferably, not more than 30 amino acid substitutions, and still even more preferably, not more than 20 amino acid substitutions. Of course, in order of ever-increasing preference, it is highly preferable for a polynucleotide which encodes the amino acid sequence of a Nodal or Lefty polypeptide to have an amino acid sequence which contains not more than 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 amino acid substitutions. Conservative substitutions are preferable.

The present invention also relates to recombinant vectors, which include the isolated nucleic acid molecules of the present invention, and to host cells containing the recombinant vectors, as well as to methods of making such vectors and host cells and for using them for production of Nodal or Lefty polypeptides or peptides by recombinant techniques.

In accordance with a further embodiment of the present invention, there is provided a process for producing such polypeptide by recombinant techniques comprising culturing recombinant prokaryotic and/or eukaryotic host cells, containing a human Nodal or Lefty nucleic acid sequence, under conditions promoting expression of said protein and subsequent recovery of said protein.

The invention further provides an isolated Nodal or Lefty polypeptide comprising an amino acid sequence selected from the group consisting of: (a) the amino acid sequence of the full-length Nodal polypeptide having the complete amino acid sequence shown in SEQ ID NO:2 (i.e., positions 1 to 283 of SEQ ID NO:2); (b) the amino acid sequence of the predicted active Nodal polypeptide having the amino acid sequence at positions 173 to 283 of SEQ ID NO:2; (c) the amino acid sequence of the Nodal polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 and/or 209135; (d) the amino acid sequence of the active domain of the Nodal polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 and/or 209135; (e) the amino acid sequence of the Lefty polypeptide having the complete amino acid sequence in SEQ ID NO:4 (i.e., positions -18 to 348 of SEQ ID NO:4); (f) the amino acid sequence of the Lefty polypeptide having the complete amino acid sequence in SEQ ID NO:4 excepting the N-terminal methionine (i.e., positions -17 to 348 of SEQ ID NO:4); (g) the amino acid sequence of the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 60 to 348 of SEQ ID NO:4; (h) the amino acid sequence of the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 118 to 348 of SEQ ID NO:4; (i) the amino acid sequence of the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 125 to 348 of SEQ ID NO:4; (j) the amino acid sequence of the Lefty polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091; (k) the amino acid sequence of the Lefty polypeptide having the complete amino acid sequence excepting the N-terminal methionine encoded by the cDNA clone contained in ATCC Deposit No. 209091, and; (l) the amino acid sequence of the active domain of the Lefty polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091.

The polypeptides of the present invention also include polypeptides having an amino acid sequence at least 80% identical, more preferably at least 90% identical, and still more preferably 95%, 96%, 97%, 98% or 99% identical to those described in (a) through (l) above, as well as polypeptides having an amino acid sequence with at least 90% similarity, and more preferably at least 95% similarity, to those above.

An additional embodiment of the invention relates to a peptide or polypeptide which comprises the amino acid sequence of an epitope-bearing portion of a Nodal or Lefty polypeptide having an amino acid sequence described in (a) through (l), above. Peptides or polypeptides having the amino acid sequence of an epitope-bearing portion of a Nodal or Lefty polypeptide of the invention include portions of such polypeptides with at least six or seven, preferably at least nine, and more preferably at least about 30 amino acids to about 50 amino acids, although epitope-bearing polypeptides of any length up to and including the entire amino acid sequence of a polypeptide of the invention described above also are included in the invention.

A further embodiment of the invention relates to a polypeptide which comprises the amino acid sequence of a Nodal or Lefty polypeptide having an amino acid sequence which contains at least one amino acid substitution, but not more than 50 amino acid substitutions, even more preferably, not more than 40 amino acid substitutions, still more preferably, not more than 30 amino acid substitutions, and still even more preferably, not more than 20 amino acid substitutions. Of course, in order of ever-increasing preference, it is highly preferable for a peptide or polypeptide to have an amino acid sequence which comprises the amino acid sequence of a TNF-gamma polypeptide, which contains at least one, but not more than 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 amino acid substitutions. In specific embodiments, the number of additions, substitutions, and/or deletions in the amino acid sequence of Figures 1A and 1B, Figures 2A and

2B, or fragments thereof (e.g., the mature form and/or other fragments described herein), is 1-5, 5-10, 5-25, 5-50, 10-50 or 50-150, conservative amino acid substitutions are preferable.

In another embodiment, the invention provides an isolated antibody that binds specifically to a Nodal and Lefty polypeptide having an amino acid sequence described in (a) through (l) above. The invention further provides methods for isolating antibodies that bind specifically to a Nodal or Lefty polypeptide having an amino acid sequence as described herein. Such antibodies are useful diagnostically or therapeutically as described below.

The invention also provides for pharmaceutical compositions comprising Nodal and Lefty polypeptides, particularly human Nodal and Lefty polypeptides, which may be employed, for instance, to treat cellular growth and differentiation disorders. Methods of treating individuals in need of Nodal and Lefty polypeptides are also provided.

The invention further provides compositions comprising a Nodal or Lefty polynucleotide or a Nodal or Lefty polypeptide for administration to cells *in vitro*, to cells *ex vivo* and to cells *in vivo*, or to a multicellular organism. In certain particularly preferred embodiments of the invention, the compositions comprise a Nodal or Lefty polynucleotide for expression of a Nodal or Lefty polypeptide in a host organism for treatment of disease. Particularly preferred in this regard is expression in a human patient for treatment of a dysfunction associated with aberrant endogenous activity of Nodal or Lefty.

The present invention also provides a screening method for identifying compounds capable of enhancing or inhibiting a biological activities of the Nodal and Lefty polypeptides, which involves contacting a receptor which is enhanced by the Nodal or Lefty polypeptides with the candidate compound in the presence of a Nodal or Lefty polypeptide, assaying receptor activation in the presence of the candidate compound and of Nodal or Lefty polypeptide, and

comparing the receptor activity to a standard level of activity, the standard being assayed when contact is made between the receptor and in the presence of the Nodal or Lefty polypeptide and the absence of the candidate compound In this assay, an increase in receptor activation over the standard indicates that the candidate compound is an agonist of Nodal or Lefty activity and a decrease in receptor activation compared to the standard indicates that the compound is an antagonist of Nodal or Lefty activity.

In another embodiment, a screening assay for agonists and antagonists is provided which involves determining the effect a candidate compound has on Nodal or Lefty binding to a receptor. In particular, the method involves contacting the receptor with a Nodal or Lefty polypeptide and a candidate compound and determining whether Nodal or Lefty polypeptide binding to the receptor is increased or decreased due to the presence of the candidate compound. In this assay, an increase in binding of Nodal or Lefty over the standard binding indicates that the candidate compound is an agonist of Nodal or Lefty binding activity and a decrease in Nodal or Lefty binding compared to the standard indicates that the compound is an antagonist of Nodal or Lefty binding activity.

It has been discovered that, by detection in the HGS EST database, Nodal is expressed not only in neutrophils, but also in testes. In addition, it has been discovered that, by detection in the HGS EST database, Lefty is expressed not only in uterine cancer, but also in colon cancer, apoptotic T-cells, fetal heart, Wilm's Tumor tissue, frontal lobe of the brain from a patient with dementia, neutrophils, salivary gland, small intestine, 7, 8, and 12 week old human embryos, frontal cortex and hypothalamus from a patient with schizophrenia, brain from a patient with Alzheimer's Disease, adipose tissue, brown fat, TNF- and LPS-induced and uninduced bone marrow stroma, activated monocytes and macrophages, rhabdomyosarcoma, cycloheximide-treated Raji cells, breast lymph nodes, hemangiopericytoma, testes, fetal epithelium (skin), and IL-5-induced

eosinophils. Therefore, nucleic acids of the invention are useful as hybridization probes for differential identification of the tissue(s) or cell type(s) present in a biological sample. Similarly, polypeptides and antibodies directed to those polypeptides are useful to provide immunological probes for differential identification of the tissue(s) or cell type(s). In addition, for a number of disorders of the above tissues or cells, particularly with regard to the regulation of cell growth and differentiation, significantly higher or lower levels of Nodal or Lefty gene expression may be detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) taken from an individual having such a disorder, relative to a "standard" Nodal or Lefty gene expression level, i.e., the Nodal and Lefty expression levels in healthy tissue from an individual not having the cell growth and differentiation disorder. Thus, the invention provides a diagnostic method useful during diagnosis of such a disorder, which involves: (a) assaying Nodal and Lefty gene expression level in cells or body fluid of an individual; (b) comparing the Nodal and Lefty gene expression levels with standard Nodal and Lefty gene expression levels, whereby an increase or decrease in the assayed Nodal and Lefty gene expression level compared to the standard expression level is indicative of disorder in the regulation of cell growth and differentiation.

An additional embodiment of the invention is related to a method for treating an individual in need of an increased level of Nodal or Lefty activity in the body comprising administering to such an individual a composition comprising a therapeutically effective amount of an isolated Nodal or Lefty polypeptide of the invention or an agonist thereof.

A still further embodiment of the invention is related to a method for treating an individual in need of a decreased level of Nodal or Lefty activity in the body comprising, administering to such an individual a composition comprising a therapeutically effective amount of a Nodal or Lefty antagonist. Preferred

antagonists for use in the present invention are Nodal- or Lefty-specific antibodies.

Brief Description of the Figures

Figures 1A and 1B show the nucleotide sequence (SEQ ID NO:1) and deduced amino acid sequence (SEQ ID NO:2) of the human Nodal homologue of the present invention.

The predicted TGF-β consensus cleavage sequences (arginine-X-X-arginine (RXXR); where X is any amino acid) of the human Nodal homologue is double underlined in Figures 1A and 1B. The TGF-β consensus cleavage sequence appears once in the amino acid sequence of Nodal. Cleavage of the precursor form of human Nodal is predicted to occur immediately after the C-terminal arginine in the abovementioned consensus sequence in the amino acid sequence of Nodal.

Potential asparagine-linked glycosylation sites are marked in Figures 1A and 1B with a bolded asparagine symbol (**N**) in the Nodal amino acid sequence and a bolded pound sign (#) above the first nucleotide encoding that asparagine residue in the Nodal nucleotide sequence. Potential N-linked glycosylation sequences are found at the following locations in the Nodal amino acid sequence: N-8 through F-11 (N-8, W-9, T-10, F-11) and N-135 through Q-138 (N-135, L-136, S-137, Q-138). A potential Protein Kinase C (PKC) phosphorylation site is also marked in Figures 1A and 1B with a bolded serine symbol (**S**) in the Nodal amino acid sequence and an asterisk (*) above the first nucleotide encoding that serine residue in the Nodal nucleotide sequence. The potential PKC phosphorylation sequence is found in the Nodal amino acid sequence from residue S-155 through residue R-157 (S-155, W-156, R-157). Potential Casein Kinase II (CK2) phosphorylation sites are also marked in Figures 1A and 1B with a bolded serine symbol (**S**) in the Nodal amino acid sequence and an asterisk

(*) above the first nucleotide encoding the appropriate serine residue in the Nodal nucleotide sequence. Potential CK2 phosphorylation sequences are found at the following locations in the Nodal amino acid sequence: S-19 through E-22 (S-19, Q-20, Q-21, E-22); S-35 through D-38 (S-35, P-36, V-37, D-38); and S-63 through E-66 (S-63, C-64, L-65, E-66). A potential myristylation site is found in the Nodal amino acid sequence in Figures 1A and 1B from residue G-6 through F-11 (G-6, Q-7, N-8, W-9, T-10, F-11). A potential amidation site is found in the Nodal amino acid sequence in Figures 1A and 1B from residue W-167 through R-170 (W-167, G-168, K-169, R-170). A TGF-beta family signature is found in the Nodal amino acid sequence in Figures 1A and 1B from residue I-201 through C-216 (I-201, I-202, Y-203, P-204, K-205, Q-206, Y-207, N-208, A-209, Y-210, R-211, C-212, E-213, G-214, E-215, C-216). This sequence is denoted in Figures 1A and 1B with a dotted underline shown under the amino acid sequence from residue I-201 through C-216.

Figures 2A and 2B show the nucleotide sequence (SEQ ID NO:3) and deduced amino acid sequence (SEQ ID NO:4) of the Lefty homologue of the present invention.

The predicted leader cleavage sequence of the human Lefty homologue of about 18 amino acids is underlined in Figure 2A. Note that the methionine residue at the beginning of the leader sequence in Figure 2A is shown in position number (positive or "+") 1, whereas the leader positions in the corresponding sequence of SEQ ID NO:2 are designated with negative position numbers. Thus, the leader sequence positions 1 to 18 in Figure 2A correspond to positions -18 to -1 in SEQ ID NO:2.

The predicted consensus sequences (arginine-X-X-arginine (RXXR); where X is any amino acid) of the human Lefty homologue is double underlined in Figures 2A and 2B. The TGF-β consensus cleavage sequence appears three times in the amino acid sequence of Lefty. Cleavage of the precursor forms of human

Lefty is predicted to occur immediately after the C-terminal arginine in the abovementioned consensus sequence in the amino acid sequence of Lefty.

A potential asparagine-linked glycosylation site is marked in Figures 2A and 2B with a bolded asparagine symbol (**N**) in the Nodal amino acid sequence and a bolded pound sign (#) above the first nucleotide encoding that asparagine residue in the Lefty nucleotide sequence. The potential N-linked glycosylation sequence is found in the Lefty amino acid sequence from residue N-158 through S-161 (N-158, R-159, T-160, S-161). A potential cAMP- and cGMP-dependent protein kinase (CPK) phosphorylation site is marked in Figures 2A and 2B with a bolded serine symbol (**S**) in the Lefty amino acid sequence and an asterisk (*) above the first nucleotide encoding that serine residue in the Lefty nucleotide sequence. The potential CPK phosphorylation sequence is found in the Lefty amino acid sequence from residue K-76 through residue S-79 (K-76, R-77, F-78, S-79). Several potential Protein Kinase C (PKC) phosphorylation sites are also marked in Figures 2A and 2B with a bolded serine or threonine symbol (**S** or **T**) in the Lefty amino acid sequence and an asterisk (*) above the first nucleotide encoding that serine or threonine residue in the Lefty nucleotide sequence. The potential PKC phosphorylation sequences are found in the Lefty amino acid sequence from residue S-81 through residue R-83 (S-81, F-82, R-83); S-137 through R-139 (S-137, P-138, R-139); S-140 through R-142 (S-140, A-141, R-142); S-157 through R-159 (S-157, N-158, R-159); T-296 through R-298 (T-296, C-297, R-298); and S-329 through K-331 (S-329, I-330, K-331). Potential Casein Kinase II (CK2) phosphorylation sites are also marked in Figures 2A and 2B with a bolded serine symbol (**S**) in the Nodal amino acid sequence and an asterisk (*) above the first nucleotide encoding the appropriate serine residue in the Lefty nucleotide sequence. Potential CK2 phosphorylation sequences are found at the following locations in the Lefty amino acid sequence: S-68 through D-71 (S-68, H-69, G-70, D-71); S-81 through E-84 (S-81, F-82,

R-83, E-84); S-161 through D-164 (S-161, L-162, I-163, D-164); S-169 through E-172 (S-169, V-170, H-171, E-172); S-319 through D-322 (S-319, E-320, T-321, D-322); and S-329 through E-332 (S-329, I-330, K-331, E-332). Several potential myristylation sites are found in the Lefty amino acid sequence in Figures 2A and 2B at the following locations: from residue G-19 through G-24 (G-19, A-20, A-21, L-22, T-23, G-24); G-156 through S-161 (G-156, S-157, N-158, R-159, T-160, S-161); G-225 through L-230 (G-225, A-226, P-227, A-228, G-229, L-230); G-260 through G-265 (G-260, T-261, R-262, C-263, C-264, R-265); and G-274 through G-279 (G-274, M-275, K-276, W-277, A-278, E-279). A potential amidation site is found in the Lefty amino acid sequence in Figures 2A and 2B from residue R-74 through R-77 (R-74, G-75, K-76, R-77). A TGF-beta family signature is found in the Lefty amino acid sequence in Figures 2A and 2B from residue V-282 through C-297 (V-282, L-283, E-284, P-285, P-286, G-287, F-288, L-289, A-290, Y-291, E-292, C-293, V-294, G-295, T-296, C-297). This sequence is denoted in Figures 2A and 2B with a dotted underline shown under the amino acid sequence from residue I-282 through C-297.

Figures 3 and 4 show the regions of identity between the amino acid sequences of the Nodal and Lefty proteins and translation product of the murine mRNAs for Nodal and Lefty, respectively, (SEQ ID NO:5 and SEQ ID NO:6, respectively), determined by the computer program Bestfit (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711) using the default parameters.

Figures 5 and 6 show computer analyses of the Nodal and Lefty amino acid sequences depicted in Figures 1A and 1B (SEQ ID NO:2) and 2A and 2B (SEQ ID NO:4), respectively. Alpha, beta, turn and coil regions; hydrophilicity and hydrophobicity; amphipathic regions; flexible regions; antigenic index and surface probability, as predicted using the default parameters of the recited

programs, are shown. In the "Antigenic Index or Jameson-Wolf" graph, the positive peaks indicate locations of the highly antigenic regions of the Nodal and Lefty proteins, i.e., regions from which epitope-bearing peptides of the invention can be obtained. Non-limiting examples of antigenic polypeptides or peptides that can be used to generate Nodal-specific antibodies include: a polypeptide comprising amino acid residues from about Lys-54 to about Asp-62, from about Val-91 to about Leu-99, from about Lys-100 to about Gln-108, from about Cys-116 to about Pro-124, from about Gln-140 to about Leu-148, from about Trp-156 to about Ser-164, from about Arg-170, to about Gln-181, from about Cys-212 to about Phe-224, from about Tyr-239, to about Thr-247, from about Pro-251, to about Met-259, and from about Asp-263, to about His-271. Non-limiting examples of antigenic polypeptides or peptides that can be used to generate Lefty-specific antibodies include: a polypeptide comprising amino acid residues from about Asp-71 to about Ser-79, from about Arg-106 to about Val-114, from about Leu-136 to about Arg-144, from about Asp-154 to about Asp-164, from about His-171 to about Asp-179, from about Gln-189 to about Leu-197, from about Pro-227 to about Glu-236, from about Gly-246 to about Glu-254, from about Pro-256 to about Gln-266, from about Cys-297 to about Ala-305, from about Ile-317 to about Pro-325, from about Ile-330 to about Val-340, and from about Val-348 to about Pro-366.

The data presented in Figures 5 and 6 are also represented in tabular form in Tables I and II, respectively. The columns are labeled with the headings "Res", "Position", and Roman Numerals I-XIV. The column headings refer to the following features of the amino acid sequence presented in Figures 5 and 6, and Tables I and II, respectively: "Res": amino acid residue of SEQ ID NO:2 or Figures 2A and 2B (which is the identical sequence shown in SEQ ID NO:4, with the exception that the residues are numbered 1-366 in Figures 2A and 2B and -18 through 348 in SEQ ID NO:4); "Position": position of the corresponding residue

within SEQ ID NO:2 or Figures 2A and 2B (which is the identical sequence shown in SEQ ID NO:4, with the exception that the residues are numbered 1-366 in Figures 2A and 2B and -18 through 348 in SEQ ID NO:4); I: Alpha, Regions - Garnier-Robson; II: Alpha, Regions - Chou-Fasman; III: Beta, Regions - Garnier-Robson; IV: Beta, Regions - Chou-Fasman; V: Turn, Regions - Garnier-Robson; VI: Turn, Regions - Chou-Fasman; VII: Coil, Regions - Garnier-Robson; VIII: Hydrophilicity Plot - Kyte-Doolittle; IX: Hydrophobicity Plot - Hopp-Woods; X: Alpha, Amphipathic Regions - Eisenberg; XI: Beta, Amphipathic Regions - Eisenberg; XII: Flexible Regions - Karplus-Schulz; XIII: Antigenic Index - Jameson-Wolf; and XIV: Surface Probability Plot - Emini.

Detailed Description

The present invention provides isolated nucleic acid molecules comprising polynucleotides encoding a Nodal or Lefty polypeptide having the amino acid sequences shown in SEQ ID NO:2 and SEQ ID NO:4, respectively, which were determined by sequencing cloned cDNAs. The nucleotide sequences shown in Figures 1A and B and 2A and B (SEQ ID NO:1 and SEQ ID NO:3, respectively) were obtained by sequencing the HNGEF08 and HUKEJ46 clones, which were deposited on June 5, 1997 at the American Type Culture Collection, 10801 University Boulevard, Manassas, Virginia 20110-2209, and given accession numbers ATCC 209092 and 209135, and 209091, respectively. The deposited clones are contained in the pBluescript SK(-) plasmid (Stratagene, La Jolla, CA).

The Nodal and Lefty proteins of the present invention share sequence homology with the translation products of the murine mRNAs for Nodal and Lefty (Figures 3 and 4). Murine Nodal is thought to be an important TGF-β superfamily member involved in mesoderm formation during gastrulation (Zhou, X., *et al., Nature* **361**:543-547 (1993)). During gastrulation, the three germ layers

of the embryo are formed and organized along the anterior-posterior body axis. In addition, ectodermal cells of the primitive streak differentiate into the mesoderm. Murine Nodal was identified in mice which were homozygously mutated in the Nodal gene. A mutation in Nodal is prenatally lethal presumably due to the resulting gross developmental abnormalities.

Murine Lefty is involved in the developmental processes which establish lateral symmetry or handedness of the maturing embryonic organism (Meno, C., *et al., Nature* **381**:151-155 (1996)). Lefty is believed to be a diffusable morphogen, the expression of which may result in the initiation of determination of symmetrical development in the mouse embryo. Lefty is transiently expressed in the left half of the gastrulating embryo just before the initiation of lateral symmetry.

Nucleic Acid Molecules

Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined using an automated DNA sequencer (such as the Model 373 from Applied Biosystems, Inc., Foster City, CA), and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of a DNA sequence determined as above. Therefore, as is known in the art for any DNA sequence determined by this automated approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion in a determined nucleotide sequence compared to the actual sequence will cause a frame shift in

translation of the nucleotide sequence such that the predicted amino acid sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

By "nucleotide sequence" of a nucleic acid molecule or polynucleotide is intended, for a DNA molecule or polynucleotide, a sequence of deoxyribonucleotides, and for an RNA molecule or polynucleotide, the corresponding sequence of ribonucleotides (A, G, C and U), where each thymidine deoxyribonucleotide (T) in the specified deoxyribonucleotide sequence is replaced by the ribonucleotide uridine (U).

Using the information provided herein, such as the nucleotide sequences in Figures 1A and B and 2A and B (SEQ ID NO:1 and SEQ ID NO:3, respectively), nucleic acid molecules of the present invention encoding a Nodal and Lefty polypeptide may be obtained using standard cloning and screening procedures, such as those for cloning cDNAs using mRNA as starting material. Illustrative of the invention, the nucleic acid molecules described in Figures 1A and B and 2A and B (SEQ ID NO:1 and SEQ ID NO:3, respectively) were discovered in cDNA libraries derived from neutrophils and uterine cancer, respectively. An additional clone of the Nodal gene was found in testis tissue. Additional clones of the Lefty gene were also identified in cDNA libraries from the following cell and tissue types: colon cancer, apoptotic T-cells, fetal heart, Wilm's Tumor tissue, frontal lobe of the brain from a patient with dementia, neutrophils, salivary gland, small intestine, 7, 8, and 12 week old human embryos, frontal cortex and hypothalamus from a patient with schizophrenia, brain from a patient with Alzheimer's Disease, adipose tissue, brown fat, TNF- and LPS-induced and uninduced bone marrow stroma, activated monocytes and macrophages, rhabdomyosarcoma, cycloheximide-treated Raji cells, breast lymph nodes, hemangiopericytoma, testes, fetal epithelium (skin), and IL-5-induced eosinophils.

Each of the determined nucleotide sequences of the Nodal and Lefty cDNAs shown in Figures 1A and B and 2A and B (SEQ ID NO:1 and SEQ ID NO:3, respectively) contains an open reading frame. The open reading frame found in Figures 1A-B encodes a protein of 283 amino acid residues, with an initiating aspartic acid codon at nucleotide positions 1-3 of the nucleotide sequence in Figure 1A (SEQ ID NO:1), and a deduced molecular weight of about 32.5 kDa. The open reading frame found in Figures 2A-B encodes a protein of 366 amino acid residues, with an initiating methionine codon at nucleotide positions 53-55 of the nucleotide sequence in Figure 2A (SEQ ID NO:3), and a deduced molecular weight of about 40.9 kDa. The amino acid sequence of the Nodal and Lefty proteins shown in SEQ ID NO:2 and SEQ ID NO:4, respectively, is about 80.9% and 82.0% identical to the murine mRNAs for Nodal and Lefty, respectively (Figures 3 and 4). The murine Nodal and Lefty genes have been described previously in the literature (Zhou, X., *et al., Nature* **361**:543-547 (1993); Bouillet, P., *et al., Dev. Biol.* **170**:420-433 (1995); Meno, C., *et al., Nature* **381**:151-155 (1996)) and can be accessed on GenBank as Accession Nos. X70514 and Z73151, respectively.

The open reading frame of the Nodal gene shares sequence homology with the translation product of the murine mRNA for Nodal; Figure 3; SEQ ID NO:3), particularly in the conserved active domain of about 110 amino acids. The open reading frame of the Lefty gene shares sequence homology with the translation product of the murine mRNA for Lefty; Figure 4; SEQ ID NO:4), particularly in the conserved active domain of about 288 amino acids. Murine Nodal is thought to be important in correct mesoderm formation in the developing mouse embryo. Murine Lefty is thought to be important in the initiation of lateral a symmetry in the developing mouse embryo. The homologies between the murine Nodal and Lefty mRNAs and the novel human homologues of Nodal and Lefty indicate that the novel human homologues of Nodal and Lefty are involved in developmental

roles as well as in the regulation of cell growth and differentiation. Further, it is likely that aberrant expression of Nodal and Lefty is a characteristic of cancer.

As members of the TGF-β superfamily, the novel human genes of the instant application also function in the regulation of immune and hematopoietic cell growth and differentiation.

As one of ordinary skill would appreciate, due to the possibilities of sequencing errors discussed above, the actual complete Nodal and Lefty polypeptides encoded by the deposited cDNAs, which comprise about 283 and 348 amino acids, respectively, may be somewhat longer or shorter. More generally, the actual open reading frame may be anywhere in the range of ±20 amino acids, more likely in the range of ±10 amino acids, of that predicted from either the codon at the N-terminus shown in Figures 1A and B and 2A and B (SEQ ID NO:1 and SEQ ID NO:3, respectively). It will further be appreciated that, depending on the analytical criteria used for identifying various functional domains, the exact "address" of the active domains of the Nodal and Lefty polypeptides may differ slightly from the predicted positions above.

Methods for predicting whether a protein has a secretory leader as well as the cleavage point for that leader sequence are known in the art and may routinely be applied to identify the leader sequence of the polynucleotides of the invention. For instance, the method of McGeoch (*Virus Res.* **3**:271-286 (1985)) uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje (*Nucleic Acids Res.* **14**:4683-4690 (1986)) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2 where +1 indicates the amino terminus of the mature protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80% (von Heinje, *supra*). However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequences of the complete Nodal and Lefty polypeptides were analyzed by a computer program "PSORT", available from Dr. Kenta Nakai of the Institute for Chemical Research, Kyoto University (Nakai, K. and Kanehisa, M. *Genomics* **14**:897-911 (1992)), which is an expert system for predicting the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated.

In one embodiment, the computation analysis above predicted a single N-terminal signal sequence within the complete amino acid sequence shown in SEQ ID NO:4. Thus, the amino acid sequence of the complete Lefty protein includes a leader sequence and a mature protein, as shown in Figures 2A and 2B and SEQ ID NO:4. The amino acid sequence of the complete Nodal protein predicts a leader sequence and a mature protein, by comparison to the full-length murine Nodal ORF as shown in Figure 3.

The present invention provides nucleic acid molecules encoding a mature form of the Lefty protein. According to the signal hypothesis, once export of the growing protein chain across the rough endoplasmic reticulum has been initiated, proteins secreted by mammalian cells have a signal or secretory leader sequence which is cleaved from the complete polypeptide to produce a secreted "mature" form of the protein. Most mammalian cells and even insect cells cleave secreted proteins with the same specificity. However, in some cases, cleavage of a secreted protein is not entirely uniform, which results in two or more mature species of the protein. Further, it has long been known that the cleavage specificity of a secreted protein is ultimately determined by the primary structure of the complete protein, that is, it is inherent in the amino acid sequence of the polypeptide. Therefore, the present invention provides a nucleotide sequence encoding the mature Lefty polypeptide having the amino acid sequence encoded by the cDNA clone contained in the host identified as ATCC Deposit No.

209091. By the "mature Lefty polypeptide having the amino acid sequence encoded by the cDNA clone in ATCC Deposit No. 209091" is meant the mature form(s) of the Lefty protein produced by expression in a mammalian cell (e.g., COS cells, as described below) of the complete open reading frame encoded by the human DNA sequence of the clone contained in the vector in the deposited host.

Nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance, cDNA and genomic DNA obtained by cloning or produced synthetically. The DNA may be double-stranded or single-stranded. Single-stranded DNA or RNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand or complementary strand.

In specific embodiments, the polynucleotides of the invention are less than 300 kb, 200 kb, 100 kb, 50 kb, 15 kb, 10 kb or 7.5 kb in length. In a further embodiment, polynucleotides of the invention comprise at least 15 contiguous nucleotides of Human Nodal or Human Lefty coding sequence, but do not comprise all or a portion of any Human Nodal or Human Lefty intron. In another embodiment, the nucleic acid comprising Human Nodal or Human Lefty coding sequence does not contain coding sequences of a genomic flanking gene (i.e., 5' or 3' to the Human Nodal or Human Lefty coding sequences in the genome).

By "isolated" nucleic acid molecule(s) is intended a nucleic acid molecule, DNA or RNA, which has been removed from its native environment For example, recombinant DNA molecules contained in a vector are considered isolated for the purposes of the present invention. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. However, a nucleic acid contained in a clone that is a member of a library (e.g., a genomic or cDNA library) that has not been isolated from other members of the

library (e.g., in the form of a homogeneous solution containing the clone and other members of the library) or which is contained on a chromosome preparation (e.g., a chromosome spread), is not "isolated" for the purposes of this invention. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Isolated nucleic acid molecules of the present invention include DNA molecules comprising an open reading frame (ORF) with an initiating codon at positions 1-3 of the nucleotide sequence shown in Figure 1A (SEQ ID NO:1) and DNA molecules comprising an open reading frame (ORF) with an initiation codon at positions 53-55 of the nucleotide sequence shown in Figure 2A (SEQ ID NO:3).

Also included are DNA molecules comprising the coding sequence for the predicted mature Lefty protein shown at positions 1-366 of SEQ ID NO:4.

In addition, isolated nucleic acid molecules of the invention include DNA molecules which comprise a sequence substantially different from those described above, but, which, due to the degeneracy of the genetic code, still encode the Nodal or Lefty proteins. Of course, the genetic code and species-specific codon preferences are well known in the art. Thus, it would be routine for one skilled in the art to generate the degenerate variants described above, for instance, to optimize codon expression for a particular host (e.g., change codons in the human mRNA to those preferred by a bacterial host such as *E. coli*).

In another embodiment, the invention provides isolated nucleic acid molecules encoding the Nodal and Lefty polypeptides having amino acid sequences encoded by the cDNA clones contained in the plasmid deposited as ATCC Deposit Nos. 209092 and 209091 on June 5, 1997 and the plasmid deposited as ATCC Deposit No. 209135 on July 2, 1997.

Preferably, these nucleic acid molecules will encode the mature polypeptides encoded by the above-described deposited cDNA clones.

The invention further provides an isolated nucleic acid molecule having the nucleotide sequence shown in Figures 1A-B (SEQ ID NO:1) and an isolated nucleic acid molecule having the nucleotide sequence shown in Figures 2A-B (SEQ ID NO:3) or the nucleotide sequences of the Nodal and Lefty cDNAs contained in the above-described deposited clones, or a nucleic acid molecule having a sequence complementary to one of the above sequences. Such isolated molecules, particularly DNA molecules, are useful as probes for gene mapping, by *in situ* hybridization with chromosomes, and for detecting expression of the Nodal and Lefty genes in human tissue, for instance, by Northern blot analysis.

The present invention is further directed to nucleic acid molecules encoding portions of the nucleotide sequences described herein as well as to fragments of the isolated nucleic acid molecules described herein. In particular, the invention provides a polynucleotide having a nucleotide sequence representing the portion of SEQ ID NO:1 which consists of positions 1-852 of SEQ ID NO:1 and a polynucleotide having a nucleotide sequence representing the portion of SEQ ID NO:3 which consists of positions 1-1153 of SEQ ID NO:3. By a fragment of an isolated nucleic acid molecule having the nucleotide sequence of the deposited cDNAs (HTLFA20, HNGEF08, and HUKEJ46), or the nucleotide sequence shown in Figures 1A and B (SEQ ID NO:1), Figures 2A and B (SEQ ID NO:3), or the complementary strand thereto, is intended fragments at least 15 nt, and more preferably at least 20 nt, still more preferably at least 25 or 30 nt, and even more preferably, at least 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 400, or 500 nt in length. These fragments have numerous uses which include, but are not limited to, diagnostic probes and primers as discussed herein. Of course, larger fragments 50-1500 nt in length are also useful according to the present invention as are fragments corresponding to most, if not all, of the nucleotide

sequence of the deposited cDNA clone HTLFA20, the deposited cDNA clone HNGEF08, the deposited cDNA clone HUKEJ46, the nucleotide sequence depicted in Figures 1A and B (SEQ ID NO:1), or the nucleotide sequence depicted in Figures 2A and B (SEQ ID NO:4). By a fragment at least 20 nt in length, for example, is intended fragments which include 20 or more contiguous bases from the nucleotide sequence of the deposited cDNA clones (HTLFA20, HNGEF08, and HUKEJ46), the nucleotide sequence as shown in Figures 1A and B (SEQ ID NO:1) or the nucleotide sequence as shown in Figures 2A and B (SEQ ID NO:4).

In a preferred embodiment, the HUKEJ46 cDNA clone in ATCC Deposit No. 209091, which encodes the Human Lefty Homologue of the present invention, contains a cDNA insert which is represented by nucleotides 1-1596 of the sequence shown in Figures 2A and 2B.

In addition, the invention provides nucleic acid molecules having nucleotide sequences related to extensive portions of SEQ ID NO:3 which have been determined from the following related cDNA clones: HUKFN65R (SEQ ID NO:7) and HUKEJ46R (SEQ ID NO:8).

Further, the invention includes a polynucleotide comprising any portion of at least about 30 nucleotides, preferably at least about 50 nucleotides, of SEQ ID NO:1 from nucleotide 1-1130. More preferably, the invention includes a polynucleotide comprising nucleotides 250-1130, 500-1130, 750-1130, 1000-1130, 1-1000, 250-1000, 500-1000, 750-1000, 1-750, 250-750, 500-750, 1-500, 250-500, and 1-250 of SEQ ID NO:1. Likewise, the invention includes a polynucleotide comprising any portion of at least about 30 nucleotides, preferably at least about 50 nucleotides, of SEQ ID NO:3 from residue 1 to 950 and 1150 to 1688. More preferably, the invention includes a polynucleotide comprising nucleotides 250-1688, 500-1688, 750-1688, 1000-1688, 1250-1688, 1500-1688, 1-1500, 250-1500, 500-1500, 750-1500, 1000-1500, 1250-1500,

1-1250, 250-1250, 500-1250, 750-1250, 1000-1250, 1-1000, 250-1000, 500-1000, 750-1000, 1-750, 250-750, 500-750, 1-500, and 250-500 of SEQ ID NO:3.

Further specific embodiments are directed to polynucleotides corresponding to nucleotides 1-125, 1-90, 1-60, 1-30, 30-125, 30-90, 30-60, 60-125, 60-90, 90-125, 310-930, 350-930, 400-930, 450-930, 500-930, 550-930, 600-930, 650-930, 700-930, 750-930, 800-930, 850-930, 900-930, 310-900, 350-900, 400-900, 450-900, 500-900, 550-900, 600-900, 650-900, 700-900, 750-900, 800-900, 850-900, 310-850, 350-850, 400-850, 450-850, 500-850, 550-850, 600-850, 650-850, 700-850, 750-850, 800-850, 310-800, 350-800, 400-800, 450-800, 500-800, 550-800, 600-800, 650-800, 700-800, 750-800, 310-750, 350-750, 400-750, 450-750, 500-750, 550-750, 600-750, 650-750, 700-750, 310-700, 350-700, 400-700, 450-700, 500-700, 550-700, 600-700, 650-700, 310-650, 350-650, 400-650, 450-650, 500-650, 550-650, 600-650, 310-600, 350-600, 400-600, 450-600, 500-600, 550-600, 310-500, 350-500, 400-500, 450-500, 310-450, 350-450, 400-450, 310-400, 350,-400, 310-350, 1050-1596, 1100-1596, 1150-1596, 1200-1596, 1250-1596, 1300-1596, 1350-1596, 1400-1596, 1450-1596, 1500-1596, 1550-1596, 1050-1550, 1100-1550, 1150-1550, 1200-1550, 1250-1550, 1300-1550, 1350-1550, 1400-1550, 1450-1550, 1500-1550, 1050-1500, 1100-1500, 1150-1500, 1200-1500, 1250-1500, 1300-1500, 1350-1500, 1400-1500, 1450-1500, 1050-1450, 1100-1450, 1150-1450, 1200-1450, 1250-1450, 1300-1450, 1350-1450, 1400-1450, 1050-1400, 1100-1400, 1150-1400, 1200-1400, 1250-1400, 1300-1400, 1350-1400, 1050-1350, 1100-1350, 1150-1350, 1200-1350, 1250-1350, 1300-1350, 1050-1300, 1100-1300, 1150-1300, 1200-1300, 1250-1300, 1050-1250, 1100-1250, 1150-1250, 1200-1250, 1050-1200, 1100-1200, 1150-1200, 1050-1150, 1100-1150, and 1050-1100 of SEQ ID NO:3.

More generally, by a fragment of an isolated nucleic acid molecule having the nucleotide sequence of the deposited cDNAs or the nucleotide sequences shown in Figures 1A and B and 2A and B (SEQ ID NO:1 and SEQ ID NO:3, respectively) is intended fragments at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 25 nt or about 30 nt, and even more preferably, at least about 40 nt or about 45 nt in length which are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments 50-300 nt in length are also useful according to the present invention as are fragments corresponding to most, if not all, of the nucleotide sequence of the deposited cDNAs or as shown in Figures 1A and B and 2A and B (SEQ ID NO:1 and SEQ ID NO:3, respectively). By a fragment at least 20 nt in length, for example, is intended fragments which include 20 or more contiguous bases from the nucleotide sequences of the deposited cDNAs or the nucleotide sequences as shown in Figures 1A and B and 2A and B (SEQ ID NO:1 and SEQ ID NO:3, respectively). By "about" in the phrase "at least about" is meant approximately and thus may refer to the identical number recited, or alternatively may differ in number by several, a few, or, alternatively, 5, 4, 3, 2 or 1 from the recited number. Preferred nucleic acid fragments of the present invention include nucleic acid molecules encoding epitope-bearing portions of the Nodal and Lefty polypeptides as identified in Figures 5 and 6 and described in more detail below.

In specific embodiments, the polynucleotide fragments of the invention encode a polypeptide which demonstrates a functional activity. By a polypeptide demonstrating "functional activity" is meant, a polypeptide capable of displaying one or more known functional activities associated with a complete, mature or TGF-β-like active forms of the Nodal or Lefty polypeptides. Such functional activities include, but are not limited to, biological activity ((e.g., the modulation of growth, development, and differentiation of a number of cell, tissue, and organ types (e.g., fibroblasts, keratinocytes, T- and B-lymphocytes,

bone, cartilage, and other connective tissues, kidney, lung, and heart)), antigenicity [ability to bind (or compete with a Nodal or Lefty polypeptide for binding) to an anti-Nodal or anti-Lefty antibody], immunogenicity (ability to generate antibody which binds to a Nodal or Lefty polypeptide), the ability to form polymers (e.g., dimers) with other Nodal or Lefty or TGF-β polypeptides, and ability to bind to a receptor or ligand for a Nodal or Lefty polypeptide. These functional activities may routinely be determined using or routinely modifying techniques known in the art, such as, for example, immunoassays, etc.

Preferred nucleic acid fragments of the present invention also include nucleic acid molecules encoding one or more of the following domains of Nodal: amino acid residues 174-283 of SEQ ID NO:2 (i.e., the TGF-β-like domain of Nodal) and amino acid residues 1-27, 30-58, 64-82, 85-110, and 130-283 of SEQ ID NO:2. Preferred nucleic acid fragments of the present invention also include nucleic acid molecules encoding one or more of the following domains of Lefty: amino acid residues 1-348 of SEQ ID NO:4 (i.e., the mature domain of Lefty), amino acid residues 60-348 of SEQ ID NO:4 (i.e., the first predicted TGF-β-like domain of Lefty), amino acid residues 118-348 of SEQ ID NO:4 (i.e., the second predicted TGF-β-like domain of Lefty), amino acid residues 125-348 of SEQ ID NO:4 (i.e., the third predicted TGF-β-like domain of Lefty), and (-15)-(-2), 3-19, 34-51, 54-72, 75-114, 117-192, 198-209, 211-286, 290-302, and 305-348 of SEQ ID NO:4.

In specific embodiments, the polynucleotide fragments of the invention encode antigenic regions. Non-limiting examples of antigenic polypeptides or peptides that can be used to generate Nodal-specific antibodies include: a polypeptide comprising amino acid residues from about Lys-54 to about Asp-62, from about Val-91 to about Leu-99, from about Lys-100 to about Gln-108, from about Cys-116 to about Pro-124, from about Gln-140 to about Leu-148, from about Trp-156 to about Ser-164, from about Arg-170, to about Gln-181, from

about Cys-212 to about Phe-224, from about Tyr-239, to about Thr-247, from about Pro-251, to about Met-259, and from about Asp-263, to about His-271. Non-limiting examples of antigenic polypeptides or peptides that can be used to generate Lefty-specific antibodies include: a polypeptide comprising amino acid residues from about Asp-71 to about Ser-79, from about Arg-106 to about Val-114, from about Leu-136 to about Arg-144, from about Asp-154 to about Asp-164, from about His-171 to about Asp-179, from about Gln-189 to about Leu-197, from about Pro-227 to about Glu-236, from about Gly-246 to about Glu-254, from about Pro-256 to about Gln-266, from about Cys-297 to about Ala-305, from about Ile-317 to about Pro-325, from about Ile-330 to about Val-340, and from about Val-348 to about Pro-366.

In additional embodiments, the polynucleotide fragments of the invention encode functional attributes of Human Nodal or Human Lefty. Preferred embodiments of the invention in this regard include fragments that comprise alpha-helix and alpha-helix forming regions ("alpha-regions"), beta-sheet and beta-sheet forming regions ("beta-regions"), turn and turn-forming regions ("turn-regions"), coil and coil-forming regions ("coil-regions"), hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions and high antigenic index regions of Human Nodal or Human Lefty.

The data representing the structural or functional attributes of Nodal and Lefty set forth in Figures 5 and 6 and/or Tables I and II, as described above, was generated using the various modules and algorithms of the DNA*STAR set on default parameters. In a preferred embodiment, the data presented in columns VIII, IX, XIII, and XIV of Tables I and II can be used to determine regions of Nodal or Lefty which exhibit a high degree of potential for antigenicity. Regions of high antigenicity are determined from the data presented in columns VIII, IX, XIII, and/or IV by choosing values which represent regions of the polypeptide

which are likely to be exposed on the surface of the polypeptide in an environment in which antigen recognition may occur in the process of initiation of an immune response.

Certain preferred regions in these regards are set out in Figures 5 and 6, but may, as shown in Tables I and II, respectively, be represented or identified by using tabular representations of the data presented in Figures 5 and 6. The DNA*STAR computer algorithm used to generate Figures 5 and 6 (set on the original default parameters) was used to present the data in Figures 5 and 6 in a tabular format (*See* Tables I and II, respectively). The tabular format of the data in Figure 5 or in Figure 6 may be used to easily determine specific boundaries of a preferred region.

The above-mentioned preferred regions set out in Figures 5 and 6 and in Tables I and II include, but are not limited to, regions of the aforementioned types identified by analysis of the amino acid sequence set out in Figures 1A and B and 2A and B. As set out in Figures 5 and 6 and in Tables I and II, such preferred regions include Garnier-Robson alpha-regions, beta-regions, turn-regions, and coil-regions, Chou-Fasman alpha-regions, beta-regions, and coil-regions, Kyte-Doolittle hydrophilic regions and hydrophobic regions, Eisenberg alpha- and beta-amphipathic regions, Karplus-Schulz flexible regions, Emini surface-forming regions and Jameson-Wolf regions of high antigenic index (generated using the amino acid sequences set out in Figures 1 and 2, and using the default parameters of the recited computer programs).

Table I

Res	Position	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Asp	1	.	.	B	.	.	.	.	-0.36	0.07	.	*	.	-0.10	0.35
Val	2	.	.	B	.	.	.	.	-0.31	-0.36	.	*	.	0.50	0.45
Ala	3	.	.	B	.	.	.	.	0.08	-0.36	.	*	.	0.50	0.35
Val	4	.	.	B	.	.	.	.	0.47	-0.39	.	*	.	0.50	0.37
Asp	5	.	.	B	.	.	.	.	0.57	0.01	.	*	F	0.05	0.79
Gly	6	.	.	.	.	T	T	.	0.26	0.29	.	*	F	0.65	0.82
Gln	7	.	.	.	.	.	T	C	0.41	0.27	.	.	F	0.60	1.60
Asn	8	.	.	.	.	T	T	.	0.41	0.41	.	.	F	0.35	0.83
Trp	9	.	.	B	.	.	T	.	0.57	0.91	.	*	.	-0.20	0.85
Thr	10	.	A	B	.	.	.	.	0.57	1.27	.	*	.	-0.60	0.42
Phe	11	.	A	B	.	.	.	.	0.21	0.87	.	*	.	-0.60	0.44
Ala	12	.	A	B	.	.	.	.	-0.09	1.26	.	*	.	-0.60	0.36
Phe	13	.	A	B	.	.	.	.	-0.79	0.73	.	*	.	-0.60	0.34
Asp	14	.	A	.	.	T	.	.	-1.31	1.03	.	*	.	-0.20	0.34
Phe	15	.	A	.	.	T	.	.	-1.30	0.93	.	*	.	-0.20	0.27
Ser	16	.	A	.	.	.	.	C	-0.60	0.81	.	*	.	-0.40	0.43
Phe	17	A	A	.	.	.	.	.	-0.01	0.43	.	*	.	-0.60	0.44
Leu	18	A	A	.	.	.	.	.	0.69	0.83	.	*	.	-0.60	0.88
Ser	19	A	A	.	.	.	.	.	0.69	0.04	.	.	F	0.00	1.14
Gln	20	A	A	.	.	.	.	.	0.58	-0.34	.	.	F	0.60	2.20
Gln	21	A	A	.	.	.	.	.	0.29	-0.44	.	.	F	0.60	2.20
Glu	22	A	A	.	.	.	.	.	0.70	-0.63	.	.	F	0.90	1.66
Asp	23	A	A	.	.	.	.	.	0.92	-0.10	.	.	F	0.60	1.01
Leu	24	A	A	.	.	.	.	.	1.22	0.00	.	.	.	-0.30	0.59
Ala	25	A	A	.	.	.	.	.	0.41	-0.40	.	*	.	0.30	0.59
Trp	26	A	A	.	.	.	.	.	0.52	0.29	.	*	.	-0.30	0.29
Ala	27	A	A	.	.	.	.	.	-0.29	0.29	.	*	.	-0.30	0.69
Glu	28	A	A	.	.	.	.	.	-0.29	0.29	.	*	.	-0.30	0.56
Leu	29	A	A	.	.	.	.	.	-0.29	0.19	.	*	.	-0.30	0.93
Arg	30	A	A	.	.	.	.	.	-0.00	-0.04	.	*	.	0.30	0.76
Leu	31	A	A	.	.	.	.	.	-0.01	-0.16	.	*	.	0.30	0.58
Gln	32	.	A	.	.	T	.	.	0.37	0.23	.	*	.	0.10	0.95
Leu	33	.	A	.	.	T	.	.	-0.49	-0.03	.	*	.	0.70	0.75
Ser	34	.	A	.	.	.	.	C	0.32	0.61	.	*	F	-0.25	0.67
Ser	35	.	.	.	.	.	T	C	-0.60	-0.07	.	*	F	1.05	0.65
Pro	36	.	.	B	.	.	T	.	0.00	0.21	.	*	F	0.25	0.65
Val	37	.	.	B	.	.	T	.	-0.31	-0.04	*	*	F	0.85	0.75
Asp	38	.	.	B	.	.	T	.	0.50	0.06	*	*	F	0.25	0.81
Leu	39	.	.	B	.	.	.	.	0.46	-0.33	.	*	F	0.65	0.91
Pro	40	.	.	B	.	.	T	.	0.46	-0.33	.	*	F	1.00	1.21
Thr	41	A	.	.	.	.	T	.	-0.14	-0.59	.	*	F	1.15	0.97
Glu	42	A	.	.	.	.	T	.	0.12	0.10	.	*	F	0.25	0.97
Gly	43	A	.	.	.	.	T	.	-0.77	-0.09	*	*	F	0.85	0.63
Ser	44	A	A	.	.	.	.	.	0.04	0.17	.	*	F	-0.15	0.31
Leu	45	A	A	.	.	.	.	.	-0.63	-0.31	.	*	.	0.30	0.31
Ala	46	A	A	.	.	.	.	.	-1.02	0.37	.	*	.	-0.30	0.22
Ile	47	A	A	.	.	.	.	.	-1.06	0.73	*	*	.	-0.60	0.14
Glu	48	A	A	.	.	.	.	.	-0.71	0.84	*	*	.	-0.60	0.23
Ile	49	A	A	.	.	.	.	.	-0.62	0.56	*	*	.	-0.60	0.40
Phe	50	.	A	B	.	.	.	.	0.23	0.49	*	*	.	-0.60	0.88
His	51	.	A	.	.	.	.	C	0.61	-0.20	*	*	.	0.65	1.01
Gln	52	.	A	.	.	.	.	C	1.50	0.23	.	*	F	0.54	2.24
Pro	53	.	A	.	.	.	.	C	1.19	-0.46	*	*	F	1.48	4.31
Lys	54	.	.	.	.	.	T	C	2.08	-0.76	.	.	F	2.52	4.58
Pro	55	.	.	.	.	.	T	C	2.78	-1.26	.	.	F	2.86	4.58
Asp	56	.	.	.	.	T	T	.	2.22	-1.26	.	.	F	3.40	5.12
Thr	57	A	.	.	.	.	T	.	1.92	-1.19	.	.	F	2.66	2.59
Glu	58	A	.	.	.	.	.	.	2.13	-0.80	.	.	F	2.12	2.24
Gln	59	A	.	.	.	.	.	.	1.79	-1.23	.	.	F	1.78	2.24
Ala	60	A	.	.	.	.	.	.	1.33	-0.84	.	.	F	1.44	2.08
Ser	61	A	.	.	.	.	T	.	0.52	-0.76	*	.	F	1.15	0.64
Asp	62	A	.	.	.	.	T	.	0.83	-0.07	*	.	F	0.85	0.31
Ser	63	A	.	.	.	.	T	.	0.94	-0.47	*	.	F	0.85	0.53
Cys	64	A	.	.	.	.	T	.	0.24	-0.97	*	.	.	1.00	0.77
Leu	65	A	A	.	.	.	.	.	0.83	-0.57	*	*	.	0.60	0.40
Glu	66	A	A	.	.	.	.	.	0.53	-0.17	*	*	.	0.30	0.52

Table I (continued)

Res	Position	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Arg	67	A	A	.	.	.	.	.	0.53	0.06	*	*	.	-0.30	0.95
Phe	68	A	A	.	.	.	.	.	0.02	-0.51	*	*	.	0.75	1.93
Gln	69	A	A	.	.	.	.	.	-0.01	-0.51	*	*	.	0.60	0.92
Met	70	A	.	.	B	.	.	.	0.49	0.27	*	*	.	-0.30	0.41
Asp	71	A	.	.	B	.	.	.	-0.37	0.76	*	*	.	-0.60	0.68
Leu	72	A	.	.	B	.	.	.	-0.79	0.61	*	*	.	-0.60	0.29
Phe	73	.	.	B	B	.	.	.	-0.90	0.70	.	*	.	-0.60	0.42
Thr	74	.	.	B	B	.	.	.	-1.20	0.77	.	.	.	-0.60	0.21
Val	75	.	.	B	B	.	.	.	-0.60	1.16	*	.	.	-0.60	0.34
Thr	76	.	.	B	B	.	.	.	-1.46	0.87	*	.	.	-0.60	0.68
Leu	77	.	.	B	B	.	.	.	-0.96	0.73	.	.	.	-0.60	0.35
Ser	78	.	.	B	B	.	.	.	-0.96	0.73	.	*	.	-0.60	0.68
Gln	79	.	.	B	B	.	.	.	-0.94	0.87	.	*	.	-0.60	0.41
Val	80	.	.	B	B	.	.	.	-0.90	0.77	.	.	.	-0.60	0.66
Thr	81	.	.	B	B	.	.	.	-0.93	0.77	.	.	.	-0.60	0.41
Phe	82	.	.	B	B	.	.	.	-0.42	0.81	*	.	.	-0.60	0.23
Ser	83	.	.	B	.	.	.	.	-0.72	0.80	.	*	.	-0.40	0.42
Leu	84	.	.	B	.	.	.	.	-1.58	0.77	.	*	.	-0.40	0.29
Gly	85	.	.	.	B	.	.	C	-1.53	0.93	.	.	.	-0.40	0.25
Ser	86	.	.	.	B	.	.	C	-1.22	0.83	.	.	.	-0.40	0.15
Met	87	.	.	B	B	.	.	.	-1.38	0.44	.	.	.	-0.60	0.32
Val	88	.	.	B	B	.	.	.	-1.39	0.40	*	.	.	-0.60	0.24
Leu	89	.	.	B	B	.	.	.	-0.47	0.46	*	.	.	-0.60	0.26
Glu	90	.	.	B	B	.	.	.	-0.33	0.07	*	.	.	-0.30	0.51
Val	91	.	.	B	B	.	.	.	-0.84	-0.11	*	.	.	0.45	1.06
Thr	92	A	.	.	B	.	.	.	-0.54	-0.07	*	.	F	0.60	1.06
Arg	93	A	.	.	.	.	T	.	0.36	-0.37	*	.	F	0.85	0.82
Pro	94	A	.	.	.	.	T	.	0.88	-0.37	*	.	F	1.00	2.21
Leu	95	A	.	.	.	.	T	.	0.07	-0.10	*	.	F	1.00	1.61
Ser	96	A	.	.	.	.	T	.	0.97	0.10	*	.	F	0.25	0.68
Lys	97	.	A	.	.	T	.	.	1.39	0.10	*	.	F	0.49	0.88
Trp	98	.	A	B	.	.	.	.	1.07	-0.33	*	.	F	1.08	2.09
Leu	99	.	A	B	.	.	.	.	0.93	-0.59	*	.	F	1.62	2.41
Lys	100	.	A	B	.	.	.	.	1.16	-0.54	*	.	F	1.86	1.19
Arg	101	.	.	.	.	.	T	C	0.64	-0.04	*	.	F	2.40	1.14
Pro	102	.	.	.	.	.	T	C	0.60	-0.27	*	.	F	2.16	1.14
Gly	103	.	.	.	.	.	T	C	0.93	-0.96	*	.	F	2.07	0.99
Ala	104	A	.	.	.	.	T	.	1.74	-0.96	*	.	F	1.78	1.01
Leu	105	A	A	.	.	.	.	.	1.10	-0.56	*	.	F	1.14	1.13
Glu	106	A	A	.	.	.	.	.	0.69	-0.37	*	.	F	0.60	1.13
Lys	107	A	A	.	.	.	.	.	1.01	-0.41	*	.	F	0.60	1.50
Gln	108	A	A	.	.	.	.	.	0.50	-0.91	*	.	F	0.90	3.57
Met	109	A	A	.	.	.	.	.	0.50	-0.96	*	.	F	0.90	1.53
Ser	110	A	A	.	.	.	.	.	0.97	-0.46	*	*	F	0.45	0.77
Arg	111	.	A	B	.	.	.	.	0.97	-0.03	*	*	.	0.30	0.44
Val	112	.	A	B	.	.	.	.	0.26	-0.43	*	*	.	0.30	0.77
Ala	113	.	A	.	.	T	.	.	-0.03	-0.47	*	.	.	0.70	0.31
Gly	114	.	A	.	.	T	.	.	0.36	0.06	*	*	.	0.35	0.17
Glu	115	.	A	.	.	T	.	.	0.77	0.49	*	*	.	0.30	0.35
Cys	116	.	A	.	.	T	.	.	0.44	-0.16	*	*	.	1.45	0.67
Trp	117	.	.	.	.	T	T	.	1.09	-0.23	*	*	.	2.25	1.05
Pro	118	.	.	.	.	T	T	.	1.37	-0.23	*	.	F	2.50	0.94
Arg	119	.	.	.	.	.	T	C	1.50	0.26	*	.	F	1.60	2.52
Pro	120	.	.	.	.	.	T	C	1.29	0.11	*	.	F	1.35	3.70
Pro	121	.	.	.	.	T	.	.	1.37	-0.37	*	.	F	1.70	3.70
Thr	122	.	.	.	.	.	.	C	1.34	-0.30	*	.	F	1.25	1.91
Pro	123	.	.	.	.	.	T	C	1.56	0.19	*	.	F	0.60	1.78
Pro	124	.	.	.	.	.	T	C	0.59	0.16	*	.	F	0.60	1.85
Ala	125	.	.	B	.	.	T	.	-0.01	0.37	.	.	F	0.25	0.95
Thr	126	.	.	B	.	.	T	.	-0.61	0.57	.	.	F	-0.05	0.51
Asn	127	.	A	B	.	.	.	.	-0.90	0.83	.	.	.	-0.60	0.27
Val	128	.	A	B	.	.	.	.	-1.50	1.01	.	.	.	-0.60	0.27
Leu	129	.	A	B	.	.	.	.	-1.53	1.20	.	.	.	-0.60	0.15
Leu	130	.	A	B	.	.	.	.	-1.24	1.47	.	.	.	-0.60	0.15
Met	131	.	A	B	.	.	.	.	-0.93	1.46	*	.	.	-0.60	0.27
Leu	132	.	A	B	.	.	.	.	-1.74	1.21	*	.	.	-0.60	0.52

Table I (continued)

Res	Position	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Tyr	133	.	.	B	.	.	T	.	-1.19	1.21	*	.	.	-0.20	0.52
Ser	134	.	.	.	.	.	T	C	-0.38	0.91	*	.	.	0.00	0.71
Asn	135	.	.	.	.	.	T	C	0.43	0.70	.	.	F	0.30	1.48
Leu	136	.	.	.	.	.	T	C	1.03	0.01	*	*	F	0.60	1.64
Ser	137	A	A	.	.	.	.	.	1.96	-0.34	*	.	F	0.60	2.12
Gln	138	A	A	.	.	.	.	.	2.20	-0.73	*	.	F	0.90	2.58
Glu	139	.	A	B	.	.	.	.	1.69	-0.73	*	.	F	0.90	5.41
Gln	140	.	A	B	.	.	.	.	1.34	-0.73	*	.	F	1.15	3.33
Arg	141	.	A	B	.	.	.	.	1.81	-0.69	*	.	F	1.40	1.90
Gln	142	.	A	B	.	.	.	.	1.81	-0.66	.	.	F	1.65	1.09
Leu	143	.	.	.	.	T	T	.	1.50	-0.27	.	.	F	2.25	0.84
Gly	144	.	.	.	.	T	T	.	0.69	-0.19	.	.	F	2.50	0.62
Gly	145	.	.	.	.	.	T	C	-0.12	0.50	.	.	F	1.15	0.30
Ser	146	.	.	.	.	.	T	C	-0.52	0.79	.	.	F	0.90	0.30
Thr	147	.	A	.	.	.	.	C	-0.52	1.01	.	.	F	0.25	0.31
Leu	148	.	A	B	.	.	.	.	-0.30	0.59	.	.	F	-0.20	0.55
Leu	149	.	A	B	.	.	.	.	0.04	0.66	.	.	.	-0.60	0.41
Trp	150	A	A	.	.	.	.	.	0.09	0.27	.	.	.	-0.30	0.50
Glu	151	A	A	.	.	.	.	.	0.09	0.17	.	*	.	-0.30	0.81
Ala	152	A	A	.	.	.	.	.	0.11	-0.13	*	*	F	0.60	1.31
Glu	153	A	.	.	.	.	T	.	1.03	0.10	*	*	F	0.40	1.31
Ser	154	A	.	.	.	.	T	.	1.26	-0.81	.	*	F	1.30	1.48
Ser	155	A	.	.	.	.	T	.	1.54	-0.31	.	*	F	1.00	1.48
Trp	156	A	.	.	.	.	T	.	1.54	-0.41	.	*	F	1.23	1.48
Arg	157	A	.	.	.	.	.	.	1.79	-0.41	.	*	.	1.11	1.92
Ala	158	A	.	.	.	.	.	.	1.79	-0.37	.	*	F	1.49	1.42
Gln	159	A	.	.	.	.	.	.	1.28	-0.36	.	*	F	1.72	2.33
Glu	160	.	.	.	.	.	.	C	1.28	-0.59	.	*	F	2.30	0.98
Gly	161	.	.	.	.	.	.	C	1.28	-0.20	.	*	F	1.92	1.30
Gln	162	.	.	.	.	.	.	C	1.17	0.21	.	*	F	0.94	0.79
Leu	163	.	.	.	.	.	.	C	1.47	-0.19	.	*	.	1.16	0.79
Ser	164	.	.	.	.	.	.	C	1.12	0.73	*	*	.	0.03	0.84
Trp	165	A	.	.	.	.	.	.	1.17	0.73	*	*	.	-0.40	0.48
Glu	166	A	.	.	.	.	.	.	1.62	0.33	*	*	.	0.35	1.16
Trp	167	A	.	.	.	.	.	.	1.59	-0.36	*	*	.	1.25	1.70
Gly	168	A	.	.	.	.	.	.	2.51	-0.24	*	*	F	1.70	2.20
Lys	169	.	.	.	.	T	.	.	2.92	-1.16	*	.	F	2.70	2.49
Arg	170	.	.	.	.	T	.	.	3.18	-1.16	*	.	F	3.00	4.64
His	171	.	.	.	.	T	.	.	3.14	-1.57	*	.	F	2.70	6.38
Arg	172	.	.	.	.	T	.	.	2.62	-1.50	*	.	F	2.40	4.34
Arg	173	.	.	.	.	T	.	.	2.76	-0.81	.	.	.	1.95	1.83
His	174	.	.	.	.	T	.	.	2.71	-0.39	.	.	.	1.69	2.08
His	175	.	.	.	.	.	.	C	2.71	-0.89	.	*	.	1.83	1.77
Leu	176	.	.	.	.	.	T	C	2.44	-0.89	.	*	.	2.37	1.77
Pro	177	.	.	.	.	T	T	.	2.33	-0.50	.	*	F	2.76	1.74
Asp	178	.	.	.	.	T	T	.	1.41	-0.60	.	*	F	3.40	2.22
Arg	179	.	.	.	.	T	T	.	0.78	-0.41	.	.	F	2.76	2.22
Ser	180	A	.	.	B	.	.	.	0.92	-0.53	.	*	F	1.77	0.77
Gln	181	A	.	.	B	.	.	.	1.78	-0.96	*	*	F	1.43	0.90
Leu	182	A	.	.	B	.	.	.	1.13	-0.96	*	*	F	1.09	0.92
Cys	183	.	.	B	B	.	.	.	1.18	-0.31	.	*	.	0.30	0.51
Arg	184	.	.	B	B	.	.	.	0.37	-0.70	.	*	.	0.60	0.59
Lys	185	.	.	B	B	.	.	.	0.67	-0.31	*	*	F	0.45	0.62
Val	186	.	.	B	B	.	.	.	-0.19	-0.60	*	*	F	0.90	2.00
Lys	187	.	.	B	B	.	.	.	0.62	-0.53	*	*	.	0.60	0.76
Phe	188	.	.	B	B	.	.	.	0.59	-0.53	.	*	.	0.60	0.63
Gln	189	.	.	B	B	.	.	.	0.48	0.26	.	*	.	-0.30	0.74
Val	190	.	.	B	B	.	.	.	-0.38	0.01	.	*	.	-0.30	0.59
Asp	191	.	.	B	B	.	.	.	-0.41	0.70	.	*	.	-0.60	0.57
Phe	192	.	.	B	B	.	.	.	-0.80	0.60	.	*	.	-0.60	0.23
Asn	193	.	.	B	B	.	.	.	-0.39	0.63	.	*	.	-0.60	0.30
Leu	194	.	.	B	B	.	.	.	-0.73	0.90	.	*	.	-0.60	0.19
Ile	195	.	.	.	B	.	.	C	-0.18	1.33	.	*	.	-0.40	0.22
Gly	196	.	.	.	B	T	.	.	-0.47	0.93	.	.	.	-0.20	0.18
Trp	197	.	.	.	.	T	T	.	-0.66	1.44	.	.	.	0.20	0.23
Gly	198	.	.	.	.	.	T	C	-1.54	1.44	.	.	.	0.00	0.23

Table I (continued)

Res	Position		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Ser	199						T	T		-0.98	1.44				0.20	0.17
Trp	200				B			T		-0.30	1.77				-0.20	0.25
Ile	201				B					0.09	1.29				-0.40	0.38
Ile	202				B					0.38	0.86				-0.40	0.57
Tyr	203				B			T		0.48	0.87				-0.20	0.95
Pro	204						T	T		0.78	0.71			F	0.50	2.11
Lys	205						T	T		0.48	0.43			F	0.50	4.85
Gln	206						T	T		1.12	0.24			F	0.80	3.13
Tyr	207						T			2.12	0.24	*			0.45	3.17
Asn	208						T	T		1.70	-0.19				1.25	3.10
Ala	209				B			T		1.91	0.39				0.37	0.96
Tyr	210				B			T		1.52	-0.01		*		1.39	1.06
Arg	211				B			T		1.52	-0.34		*		1.51	0.65
Cys	212				B					1.10	-0.74	*	*		2.03	1.12
Glu	213						T			0.89	-0.67	*	*	F	2.70	0.38
Gly	214						T			1.48	-1.00	*	*	F	2.43	0.30
Glu	215						T			1.51	-0.60	*	*	F	2.16	0.91
Cys	216							T	C	0.54	-0.74	*	*	F	2.15	0.81
Pro	217							T	C	0.87	-0.10			F	1.84	0.61
Asn	218							T	C	0.87	-0.10			F	1.83	0.35
Pro	219							T	C	1.21	-0.10	*		F	2.24	1.12
Val	220								C	0.51	-0.67	*		F	2.60	1.26
Gly	221	A								1.14	-0.31	*		F	1.69	0.68
Glu	222	A								1.14	-0.21	*		F	1.43	0.60
Glu	223	A								0.83	-0.21	*		F	1.42	1.24
Phe	224	A								1.04	-0.37			F	1.26	1.81
His	225	A						T		1.87	-0.40			F	1.30	1.68
Pro	226	A						T		1.62	0.10			F	0.80	1.32
Thr	227						T	T		1.38	0.60			F	1.00	1.54
Asn	228	A						T		0.49	0.57		*		0.35	1.77
His	229	A			B					1.19	0.76				-0.30	0.80
Ala	230	A			B					0.92	0.73				-0.40	0.96
Tyr	231	A			B					0.32	0.63				-0.50	0.80
Ile	232			B	B					-0.18	0.91	*	*		-0.60	0.49
Gln	233			B	B					-0.13	1.10	*			-0.60	0.40
Ser	234			B	B					0.01	0.60	*			-0.60	0.51
Leu	235			B	B					0.36	-0.16	*		F	0.60	1.42
Leu	236			B	B					0.60	-0.09	*		F	0.60	1.28
Lys	237				B		T			1.28	-0.09	*		F	1.00	1.66
Arg	238						T			1.24	-0.04			F	1.20	3.11
Tyr	239			B						1.66	-0.23			F	1.08	5.13
Gln	240			B				T		1.61	-0.91			F	1.86	5.02
Pro	241			B				T		2.21	-0.27			F	1.84	1.90
His	242						T	T		1.87	0.16				1.77	1.88
Arg	243						T	T		1.44	-0.21			F	2.80	1.45
Val	244			B						1.02	-0.13	*		F	1.92	1.36
Pro	245						T			0.36	0.01	*		F	1.29	0.53
Ser	246						T	T		-0.02	0.09	*	*	F	1.21	0.15
Thr	247						T	T		-0.20	0.59	*	*	F	0.63	0.20
Cys	248			B				T		-1.17	0.37	*	*		0.10	0.20
Cys	249			B				T		-0.27	0.59		*		-0.20	0.11
Ala	250			B						-0.37	0.20		*		0.06	0.15
Pro	251			B						-0.02	0.20		*		0.22	0.41
Val	252			B						0.08	-0.37		*	F	1.28	1.53
Lys	253			B						-0.07	-0.51		*	F	1.74	2.35
Thr	254			B						0.30	-0.33		*	F	1.60	1.25
Lys	255			B						0.29	-0.37		*	F	1.44	2.26
Pro	256			B						-0.31	-0.40			F	1.28	1.12
Leu	257		A	B	B					0.30	0.29		*		0.02	0.64
Ser	258		A	B	B					-0.60	0.56				-0.44	0.50
Met	259		A	B	B					-0.29	1.20				-0.60	0.24
Leu	260		A	B	B					-0.33	0.77				-0.43	0.49
Tyr	261			B	B					-0.47	0.49				-0.26	0.58
Val	262			B				T		0.46	0.53				0.31	0.58
Asp	263			B				T		-0.10	-0.09			F	1.68	1.39
Asn	264			B				T		-0.31	-0.13		*	F	1.70	0.66

Table I (continued)

Res	Position	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Gly	265	A	.	.	.	.	T	.	-0.31	-0.20	*	*	F	1.53	0.73
Arg	266	A	A	.	.	.	.	.	-0.07	-0.16	*	*	F	0.96	0.36
Val	267	A	A	.	.	.	.	.	0.76	-0.16	*	*	.	0.64	0.37
Leu	268	A	A	.	.	.	.	.	0.72	-0.06	*	*	.	0.47	0.52
Leu	269	A	A	.	.	.	.	.	0.77	0.01	*	*	.	-0.30	0.36
Asp	270	A	A	.	.	.	.	.	1.11	0.01	*	*	.	-0.30	0.96
His	271	A	A	.	.	.	.	.	0.40	-0.63	*	*	.	0.75	1.95
His	272	A	A	.	.	.	.	.	0.37	-0.70	.	.	.	0.75	2.34
Lys	273	A	A	.	.	.	.	.	0.32	-0.70	*	.	.	0.60	0.98
Asp	274	A	A	.	.	.	.	.	1.13	-0.06	.	.	.	0.30	0.54
Met	275	A	A	.	.	.	.	.	1.13	-0.56	.	.	.	0.60	0.68
Ile	276	A	A	.	.	.	.	.	0.50	-1.06	.	.	.	0.60	0.59
Val	277	A	A	.	.	.	.	.	0.19	-0.49	.	.	.	0.30	0.19
Glu	278	A	A	.	.	.	.	.	-0.52	-0.06	.	.	.	0.30	0.19
Glu	279	A	A	.	.	.	.	.	-1.33	-0.10	*	.	.	0.30	0.15
Cys	280	A	.	.	.	.	T	.	-1.12	-0.10	.	.	.	0.70	0.16
Gly	281	A	.	.	.	.	T	.	-0.62	-0.31	.	.	.	0.70	0.12
Cys	282	A	.	.	.	.	T	.	-0.16	0.11	.	.	.	0.10	0.09
Leu	283	A	.	.	.	.	T	.	-0.54	0.54	.	.	.	-0.20	0.21

Table II

Res	Position	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Met	1	.	.	B	.	.	.	.	0.03	0.41	.	.	.	-0.40	0.82
Gln	2	.	.	B	.	.	T	.	-0.39	0.90	.	.	.	-0.20	0.67
Pro	3	.	.	B	.	.	T	.	-0.67	1.16	.	.	.	-0.20	0.43
Leu	4	.	.	.	.	T	T	.	-0.57	1.30	.	.	.	0.20	0.24
Trp	5	A	.	.	.	.	T	.	-0.77	1.60	.	.	.	-0.20	0.14
Leu	6	.	A	B	.	.	.	.	-0.98	1.70	.	.	.	-0.60	0.09
Cys	7	.	A	B	.	.	.	.	-1.27	1.96	.	.	.	-0.60	0.09
Trp	8	A	A	.	.	.	.	.	-1.91	2.19	.	.	.	-0.60	0.09
Ala	9	.	A	B	.	.	.	.	-1.91	1.91	.	.	.	-0.60	0.08
Leu	10	.	A	B	.	.	.	.	-1.83	1.91	.	.	.	-0.60	0.13
Trp	11	.	A	B	.	.	.	.	-1.83	1.77	.	.	.	-0.60	0.19
Val	12	.	A	B	.	.	.	.	-1.76	1.54	.	.	.	-0.60	0.16
Leu	13	.	A	B	.	.	.	.	-1.77	1.54	.	.	.	-0.60	0.19
Pro	14	.	.	B	.	.	.	.	-1.39	1.24	.	.	.	-0.40	0.24
Leu	15	.	.	.	.	T	.	.	-0.92	0.76	.	.	.	0.00	0.50
Ala	16	.	.	.	.	.	.	C	-1.22	0.54	.	.	.	-0.20	0.61
Ser	17	.	.	.	.	.	T	C	-0.96	0.36	.	.	F	0.45	0.40
Pro	18	.	.	.	.	.	T	C	-0.96	0.43	.	.	F	0.15	0.48
Gly	19	.	.	.	.	.	T	C	-1.06	0.43	.	.	.	0.00	0.40
Ala	20	A	.	.	.	.	T	.	-0.59	0.41	.	.	.	-0.20	0.43
Ala	21	A	A	.	.	.	.	.	-0.00	0.46	.	.	.	-0.60	0.27
Leu	22	.	A	B	.	.	.	.	0.30	0.03	.	.	.	-0.30	0.48
Thr	23	.	A	B	.	.	.	.	-0.30	0.00	.	.	F	-0.15	0.82
Gly	24	A	A	.	.	.	.	.	-0.77	0.19	.	.	F	-0.15	0.67
Glu	25	A	A	.	.	.	.	.	-0.52	0.37	.	.	F	-0.15	0.67
Gln	26	A	A	.	.	.	.	.	-0.23	0.11	.	.	F	-0.15	0.46
Leu	27	A	A	.	.	.	.	.	-0.23	0.01	.	.	F	-0.15	0.62
Leu	28	A	A	.	.	.	.	.	-0.73	0.27	*	.	F	-0.15	0.30
Gly	29	A	A	.	.	.	.	.	-0.28	0.96	*	.	F	-0.45	0.14
Ser	30	A	A	.	.	.	.	.	-0.28	0.56	*	.	F	-0.45	0.33
Leu	31	A	A	.	.	.	.	.	-1.09	0.27	*	.	F	-0.30	0.70
Leu	32	A	A	.	.	.	.	.	-0.28	0.27	*	.	.	-0.30	0.58
Arg	33	A	A	.	.	.	.	.	-0.28	0.24	*	*	.	-0.30	0.76
Gln	34	A	A	.	.	.	.	.	0.11	0.54	.	.	.	-0.60	0.76
Leu	35	A	A	.	.	.	.	.	0.41	-0.14	.	.	.	0.45	1.83
Gln	36	.	A	B	.	.	.	.	0.37	-0.83	.	.	.	0.75	1.62
Leu	37	.	A	B	.	.	.	.	0.97	-0.19	.	.	.	0.30	0.69
Lys	38	.	A	B	.	.	.	.	0.54	-0.16	.	.	F	0.60	1.30
Glu	39	.	A	B	.	.	.	.	-0.27	-0.36	.	*	F	0.60	1.08
Val	40	.	A	B	.	.	.	.	0.54	-0.07	*	*	F	0.60	1.08
Pro	41	.	A	B	.	.	.	.	0.66	-0.76	*	.	F	0.75	0.91
Thr	42	A	A	.	.	.	.	.	0.88	-0.76	*	.	F	0.90	1.02
Leu	43	A	A	.	.	.	.	.	0.83	-0.26	*	*	F	0.60	1.39
Asp	44	A	A	.	.	.	.	.	0.23	-0.90	*	*	F	0.90	1.51
Arg	45	A	A	.	.	.	.	.	1.09	-0.71	*	*	F	0.90	1.03
Ala	46	A	A	.	.	.	.	.	1.30	-1.20	.	.	F	0.90	2.17
Asp	47	A	A	.	.	.	.	.	0.80	-1.89	.	.	.	0.75	2.25
Met	48	A	A	.	.	.	.	.	0.76	-1.20	.	.	.	0.60	0.95
Glu	49	A	A	.	.	.	.	.	-0.13	-0.56	.	*	.	0.60	0.70
Glu	50	A	A	.	B	.	.	.	-0.46	-0.37	.	*	.	0.30	0.29
Leu	51	A	A	.	B	.	.	.	-0.18	0.06	.	.	.	-0.30	0.46
Val	52	A	A	.	B	.	.	.	-0.21	-0.07	.	.	.	0.30	0.38
Ile	53	A	A	.	B	.	.	.	-0.47	0.43	.	*	.	-0.60	0.30
Pro	54	A	A	.	B	.	.	.	-0.36	1.07	.	*	.	-0.60	0.27
Thr	55	A	.	.	B	.	.	.	-0.94	0.39	.	*	.	-0.30	0.71
His	56	A	A	.	B	.	.	.	-0.13	0.24	.	*	.	-0.15	1.02
Val	57	A	A	.	B	.	.	.	0.48	-0.04	.	*	.	0.45	1.14
Arg	58	.	A	B	B	.	.	.	0.51	0.29	.	*	.	-0.15	1.24
Ala	59	.	A	B	B	.	.	.	0.13	0.44	.	*	.	-0.60	0.68
Gln	60	.	A	B	B	.	.	.	-0.37	0.44	.	*	.	-0.60	0.92
Tyr	61	.	A	B	B	.	.	.	-1.14	0.49	.	*	.	-0.60	0.39
Val	62	.	A	B	B	.	.	.	-0.29	1.17	.	*	.	-0.60	0.32
Ala	63	.	A	B	B	.	.	.	-0.29	1.07	.	*	.	-0.60	0.32
Leu	64	.	A	B	B	.	.	.	-0.00	0.67	*	.	.	-0.60	0.40
Leu	65	.	A	B	B	.	.	.	-0.03	0.30	*	.	.	0.04	0.72
Gln	66	.	A	B	B	.	.	.	-0.13	0.16	*	.	.	0.38	0.96
Arg	67	.	A	B	B	.	.	.	0.72	0.09	.	.	F	1.02	1.16

Table II (continued)

Res Position		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Ser	68	.	A	.	B	T	.	.	1.42	-0.60	.	.	F	2.66	2.34
His	69	.	.	.	.	T	T	.	1.93	-1.29	*	*	F	3.40	2.65
Gly	70	.	.	.	.	T	T	.	2.86	-1.30	*	*	F	3.06	1.81
Asp	71	.	.	.	.	T	T	.	2.51	-1.30	.	*	F	3.06	2.65
Arg	72	.	.	.	.	T	T	.	2.44	-1.26	.	.	F	3.06	1.93
Ser	73	.	.	.	.	T	T	.	2.86	-1.76	.	.	F	3.06	3.90
Arg	74	.	.	.	.	T	T	.	2.19	-2.19	.	.	F	3.06	4.57
Gly	75	.	.	.	.	T	T	.	2.23	-1.40	*	.	F	3.40	2.02
Lys	76	.	.	.	.	T	T	.	2.23	-1.01	*	*	F	3.06	2.02
Arg	77	.	.	.	.	T	.	.	1.82	-1.00	*	*	F	2.72	1.79
Phe	78	.	.	B	.	.	.	.	1.42	-0.61	*	*	F	2.18	2.42
Ser	79	.	.	B	.	.	T	.	1.42	-0.26	*	*	F	1.94	1.05
Gln	80	.	.	B	.	.	T	.	1.77	-0.26	*	*	F	1.80	1.05
Ser	81	.	.	B	.	.	T	.	0.87	-0.26	*	*	F	2.00	2.09
Phe	82	.	.	B	.	.	T	.	0.17	-0.40	*	*	F	1.80	1.16
Arg	83	.	A	B	.	.	.	.	0.52	-0.29	*	*	F	1.05	0.68
Glu	84	A	A	.	.	.	.	.	0.93	-0.26	*	*	.	0.70	0.50
Val	85	A	A	.	.	.	.	.	0.23	-0.64	*	.	.	0.95	1.13
Ala	86	A	A	.	.	.	.	.	-0.28	-0.64	*	*	.	0.60	0.50
Gly	87	A	A	.	.	.	.	.	-0.17	0.04	*	.	.	-0.30	0.24
Arg	88	A	A	.	.	.	.	.	-1.09	0.54	*	.	.	-0.60	0.32
Phe	89	A	A	.	.	.	.	.	-1.09	0.59	*	*	.	-0.60	0.26
Leu	90	A	A	.	.	.	.	.	-0.82	0.09	.	*	.	-0.30	0.46
Ala	91	A	A	.	.	.	.	.	-0.53	0.16	*	.	.	-0.30	0.24
Leu	92	A	A	.	.	.	.	.	-0.50	0.54	.	.	.	-0.60	0.37
Glu	93	A	A	.	.	.	.	.	-0.64	0.24	.	*	.	-0.30	0.65
Ala	94	A	A	.	.	.	.	.	-0.76	0.06	.	.	.	-0.30	0.87
Ser	95	A	.	.	B	.	.	.	-0.76	0.24	.	*	F	-0.15	0.87
Thr	96	A	.	.	B	.	.	.	-1.02	0.24	.	.	.	-0.30	0.41
His	97	A	.	.	B	.	.	.	-0.91	0.89	.	*	.	-0.60	0.30
Leu	98	A	.	.	B	.	.	.	-1.26	1.17	.	.	.	-0.60	0.20
Leu	99	A	.	.	B	.	.	.	-1.27	1.21	.	.	.	-0.60	0.13
Val	100	A	.	.	B	.	.	.	-0.97	1.34	.	.	.	-0.60	0.10
Phe	101	.	.	B	B	.	.	.	-0.66	0.84	.	.	.	-0.60	0.21
Gly	102	.	.	B	B	.	.	.	-0.51	0.56	.	*	.	-0.60	0.43
Met	103	.	A	B	.	.	.	.	-0.51	-0.13	.	*	.	0.45	1.14
Glu	104	.	A	B	.	.	.	.	0.09	-0.09	.	*	F	0.60	1.09
Gln	105	.	A	B	.	.	.	.	0.73	-0.44	*	*	F	0.90	1.70
Arg	106	.	A	.	.	.	.	C	1.43	-0.44	.	*	F	1.40	2.66
Leu	107	.	A	.	.	.	.	C	1.48	-0.66	.	*	F	2.00	2.47
Pro	108	.	.	.	.	.	T	C	2.08	-0.27	.	*	F	2.40	1.91
Pro	109	.	.	.	.	.	T	C	1.27	-0.67	.	*	F	3.00	1.69
Asn	110	.	.	.	.	.	T	C	0.41	0.01	.	*	F	1.80	1.69
Ser	111	.	.	.	.	.	T	C	0.30	-0.03	.	*	F	1.95	0.81
Glu	112	A	A	.	.	.	.	.	0.52	-0.06	*	.	F	1.05	0.91
Leu	113	A	A	.	.	.	.	.	-0.12	0.01	.	.	.	0.00	0.57
Val	114	A	A	.	.	.	.	.	-0.72	0.26	*	*	.	-0.30	0.32
Gln	115	A	A	.	.	.	.	.	-0.61	0.56	*	*	.	-0.60	0.15
Ala	116	A	A	.	.	.	.	.	-1.12	0.56	*	*	.	-0.60	0.36
Val	117	A	A	.	.	.	.	.	-1.82	0.56	*	.	.	-0.60	0.40
Leu	118	.	A	B	.	.	.	.	-1.01	0.70	*	*	.	-0.60	0.20
Arg	119	.	A	B	.	.	.	.	-0.16	0.70	*	*	.	-0.60	0.34
Leu	120	.	A	B	.	.	.	.	-0.37	0.20	*	*	.	-0.30	0.79
Phe	121	.	A	B	.	.	.	.	-0.63	-0.01	*	.	.	0.45	1.49
Gln	122	.	A	B	.	.	.	.	0.01	-0.06	*	.	F	0.45	0.56
Glu	123	.	A	.	.	.	.	C	0.87	0.37	*	*	F	0.20	1.06
Pro	124	A	A	.	.	.	.	.	0.17	-0.31	*	.	F	0.60	2.44
Val	125	A	A	.	.	.	.	.	0.39	-0.60	*	.	F	0.90	1.42
Pro	126	A	A	.	.	.	.	.	0.28	-0.50	*	.	F	0.45	0.83
Lys	127	A	A	.	.	.	.	.	0.24	0.19	.	.	F	-0.15	0.44
Ala	128	A	A	.	.	.	.	.	0.36	0.26	.	.	.	-0.30	0.81
Ala	129	A	A	.	.	.	.	.	0.53	-0.39	.	.	.	0.45	1.03
Leu	130	A	A	.	.	.	.	.	1.04	-0.31	*	.	.	0.30	0.70
His	131	A	.	.	.	.	T	.	1.37	0.11	*	*	.	0.10	0.69
Arg	132	.	.	B	.	.	T	.	0.51	-0.39	*	*	.	0.85	1.33
His	133	.	.	.	.	T	T	.	0.80	-0.20	*	*	.	1.25	1.33

Table II (continued)

Res	Position	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Gly	134	.	.	.	.	T	T	.	1.18	-0.50	*	*	.	1.25	1.31
Arg	135	.	.	.	.	T	.	.	2.10	-0.57	*	*	F	1.84	1.03
Leu	136	.	.	.	.	.	.	C	1.83	-0.57	*	*	F	1.98	1.49
Ser	137	.	.	.	.	.	T	C	1.13	-0.69	*	*	F	2.52	2.01
Pro	138	.	.	.	.	.	T	C	1.28	-0.61	*	*	F	2.86	1.04
Arg	139	.	.	.	.	T	T	.	1.03	-0.61	*	*	F	3.40	2.47
Ser	140	.	.	.	.	.	T	C	1.03	-0.80	*	*	F	2.86	1.86
Ala	141	.	.	B	.	.	.	.	0.99	-1.19	.	*	F	2.12	2.36
Arg	142	.	.	B	B	.	.	.	0.98	-0.97	.	*	.	1.28	0.89
Ala	143	.	.	B	B	.	.	.	0.33	-0.49	.	*	.	0.64	0.96
Arg	144	.	.	B	B	.	.	.	0.22	-0.23	.	*	.	0.30	0.71
Val	145	.	.	B	B	.	.	.	0.23	-0.73	.	*	.	0.60	0.62
Thr	146	.	.	B	B	.	.	.	0.01	0.19	*	*	.	-0.30	0.65
Val	147	.	.	B	B	.	.	.	0.01	0.37	*	*	.	-0.30	0.27
Glu	148	.	.	B	B	.	.	.	-0.26	0.37	*	*	.	-0.30	0.72
Trp	149	.	.	B	B	.	.	.	-0.26	0.37	*	*	.	-0.30	0.37
Leu	150	.	.	B	B	.	.	.	0.60	-0.11	.	*	.	0.64	0.98
Arg	151	.	.	B	B	.	.	.	0.91	-0.76	.	*	.	1.28	0.95
Val	152	.	.	B	B	.	.	.	1.42	-0.76	.	*	.	1.77	1.50
Arg	153	.	.	.	B	T	.	.	1.12	-1.24	*	*	F	2.66	1.80
Asp	154	.	.	.	.	T	T	.	1.41	-1.54	*	*	F	3.40	1.23
Asp	155	.	.	.	.	T	T	.	2.33	-1.14	*	*	F	3.06	2.67
Gly	156	.	.	.	.	T	T	.	1.91	-1.79	.	*	F	2.72	2.67
Ser	157	.	.	.	.	.	T	C	2.47	-1.30	.	*	F	2.35	2.31
Asn	158	.	.	.	.	.	T	C	1.54	-0.91	.	*	F	2.18	1.85
Arg	159	.	.	B	.	.	T	.	0.66	-0.23	.	.	F	1.51	1.54
Thr	160	.	.	B	.	.	T	.	0.66	0.03	.	.	F	0.93	0.81
Ser	161	.	.	B	.	.	T	.	0.70	-0.36	.	*	F	1.70	0.84
Leu	162	.	.	B	.	.	.	.	1.11	-0.37	.	*	F	1.33	0.57
Ile	163	.	.	B	.	.	.	.	0.30	-0.37	*	.	F	1.16	0.78
Asp	164	.	.	B	.	.	T	.	-0.67	-0.17	*	*	F	1.19	0.48
Ser	165	.	.	B	.	.	T	.	-0.66	0.09	.	.	F	0.42	0.43
Arg	166	.	.	B	.	.	T	.	-1.21	-0.21	.	.	F	0.85	0.82
Leu	167	.	.	B	.	.	T	.	-0.43	-0.26	.	.	.	0.70	0.37
Val	168	.	.	B	.	.	.	.	0.46	0.24	*	*	.	-0.10	0.37
Ser	169	.	.	B	.	.	.	.	0.16	-0.14	.	.	.	0.50	0.33
Val	170	.	.	B	.	.	.	.	0.11	0.24	*	.	.	0.18	0.53
His	171	.	.	B	.	.	.	.	-0.29	-0.01	*	.	.	1.06	0.71
Glu	172	A	.	.	.	.	T	.	0.57	0.26	*	.	F	1.09	0.56
Ser	173	A	.	.	.	.	T	.	0.83	-0.13	*	.	F	2.12	1.51
Gly	174	.	.	.	.	T	T	.	0.43	-0.27	*	.	F	2.80	1.12
Trp	175	A	.	.	.	.	T	.	1.29	0.01	*	.	F	1.37	0.56
Lys	176	A	A	.	.	.	.	.	0.47	0.01	*	.	.	0.54	0.70
Ala	177	A	A	.	.	.	.	.	0.16	0.27	*	.	.	0.26	0.52
Phe	178	A	A	.	.	.	.	.	0.46	0.33	.	.	.	-0.02	0.72
Asp	179	A	A	.	.	.	.	.	0.21	-0.59	*	.	.	0.60	0.62
Val	180	A	A	.	.	.	.	.	-0.36	-0.09	.	.	.	0.30	0.62
Thr	181	A	A	.	.	.	.	.	-0.40	0.06	.	*	.	-0.30	0.53
Glu	182	A	A	.	.	.	.	.	-0.51	-0.33	*	*	.	0.30	0.51
Ala	183	A	A	.	.	.	.	.	-0.10	0.46	.	*	.	-0.60	0.60
Val	184	A	A	.	.	.	.	.	-0.10	0.73	*	.	.	-0.60	0.44
Asn	185	A	A	.	.	.	.	.	0.76	0.64	*	.	.	-0.60	0.44
Phe	186	A	A	.	.	.	.	.	0.26	1.04	*	.	.	-0.60	0.75
Trp	187	A	A	.	.	.	.	.	-0.04	1.23	*	.	.	-0.60	0.83
Gln	188	A	A	.	.	.	.	.	0.66	0.97	*	.	.	-0.60	0.69
Gln	189	.	A	.	.	T	.	.	1.30	0.57	*	*	.	0.29	1.56
Leu	190	.	A	.	.	T	.	.	1.41	0.21	*	*	F	1.08	2.30
Ser	191	.	A	.	.	.	.	C	2.11	-0.70	*	.	F	2.12	2.60
Arg	192	.	.	.	.	.	T	C	2.19	-0.70	*	*	F	2.86	2.60
Pro	193	.	.	.	.	T	T	.	1.38	-0.67	*	.	F	3.40	4.88
Arg	194	.	.	.	.	T	T	.	0.57	-0.67	.	*	F	3.06	3.00
Gln	195	.	.	B	.	.	T	.	0.57	-0.37	.	*	F	2.02	1.26
Pro	196	.	A	B	.	.	.	.	0.87	0.31	.	*	F	0.53	0.67
Leu	197	.	A	B	.	.	.	.	-0.10	0.29	.	*	F	0.19	0.60
Leu	198	.	A	B	.	.	.	.	-0.19	0.93	.	*	.	-0.60	0.26
Leu	199	.	A	B	.	.	.	.	-1.16	0.91	.	*	.	-0.60	0.22

Table II (continued)

Res	Position	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Gln	200	.	A	B	.	.	.	.	-1.16	1.13	.	.	.	-0.60	0.20
Val	201	.	A	B	.	.	.	.	-0.83	0.84	.	*	.	-0.60	0.42
Ser	202	.	.	B	B	.	.	.	-0.02	0.16	.	*	.	-0.30	0.99
Val	203	.	A	B	B	.	.	.	0.76	-0.53	.	.	.	0.60	0.99
Gln	204	.	A	B	B	.	.	.	0.76	-0.43	.	*	F	0.60	1.82
Arg	205	.	A	B	B	.	.	.	0.41	-0.39	.	.	F	0.60	1.12
Glu	206	.	A	B	B	.	.	.	1.06	-0.34	.	.	F	0.60	1.50
His	207	.	A	B	.	.	.	.	0.54	-0.56	.	.	F	0.90	1.34
Leu	208	.	A	.	.	.	.	C	0.81	-0.27	.	.	F	0.65	0.56
Gly	209	.	A	.	.	.	.	C	0.51	0.23	.	.	F	0.05	0.33
Pro	210	.	.	.	.	.	.	C	0.06	0.61	*	.	F	-0.05	0.32
Leu	211	A	.	.	.	.	.	.	-0.53	0.54	*	.	F	-0.25	0.39
Ala	212	A	.	.	.	.	T	.	-0.53	0.36	*	.	F	0.25	0.40
Ser	213	A	.	.	.	.	T	.	0.32	0.43	*	.	F	-0.05	0.35
Gly	214	A	.	.	.	.	T	.	-0.14	-0.00	*	.	.	0.70	0.84
Ala	215	A	.	.	.	.	T	.	-0.79	-0.00	*	.	.	0.70	0.69
His	216	A	A	.	.	.	.	.	0.13	0.14	*	.	.	-0.30	0.38
Lys	217	A	A	.	.	.	.	.	0.02	-0.24	*	.	.	0.30	0.76
Leu	218	.	A	B	.	.	.	.	-0.27	0.11	*	.	.	-0.30	0.65
Val	219	.	A	B	.	.	.	.	-0.22	0.11	*	.	.	-0.30	0.48
Arg	220	.	A	B	.	.	.	.	0.37	-0.00	*	*	.	0.30	0.32
Phe	221	.	A	B	.	.	.	.	0.06	0.40	*	.	.	-0.30	0.68
Ala	222	.	A	B	.	.	.	.	-0.58	0.14	*	.	.	-0.30	0.90
Ser	223	.	.	.	.	.	T	C	0.02	-0.00	*	*	F	1.05	0.47
Gln	224	.	.	.	.	T	T	.	0.29	0.43	*	*	F	0.35	0.83
Gly	225	.	.	.	.	.	T	C	-0.17	0.14	*	*	F	0.45	0.83
Ala	226	.	.	.	.	.	T	C	-0.28	0.07	.	.	F	0.66	0.61
Pro	227	.	.	.	.	.	T	C	-0.03	0.37	.	.	F	0.87	0.29
Ala	228	.	.	.	.	.	T	C	0.27	0.40	.	.	.	0.93	0.29
Gly	229	.	.	.	.	.	T	C	0.06	-0.03	.	.	.	1.74	0.50
Leu	230	.	.	.	.	.	T	C	0.40	-0.10	.	.	F	2.10	0.50
Gly	231	.	.	.	.	.	.	C	0.18	-0.13	.	*	F	1.69	0.86
Glu	232	.	A	.	.	.	.	C	0.39	0.06	.	*	F	0.68	0.72
Pro	233	A	A	.	.	.	.	.	0.17	-0.37	.	*	F	1.02	1.50
Gln	234	A	A	.	.	.	.	.	0.48	-0.37	.	*	F	0.81	1.25
Leu	235	A	A	.	.	.	.	.	0.98	-0.30	.	*	.	0.30	0.98
Glu	236	A	A	.	.	.	.	.	0.51	0.19	.	*	.	-0.30	0.92
Leu	237	A	A	.	.	.	.	.	0.51	0.44	.	*	.	-0.60	0.44
His	238	A	A	.	.	.	.	.	-0.09	0.04	.	*	.	-0.30	0.89
Thr	239	A	A	.	.	.	.	.	-0.43	0.04	.	.	.	-0.30	0.42
Leu	240	.	A	B	.	.	.	.	0.38	0.47	.	.	.	-0.60	0.51
Asp	241	.	A	B	.	.	.	.	0.13	-0.21	.	.	.	0.30	0.62
Leu	242	.	A	B	.	.	.	.	0.60	0.04	.	.	.	-0.30	0.67
Gly	243	.	.	.	.	T	T	.	0.04	-0.01	*	.	F	1.25	0.81
Asp	244	.	.	.	.	T	T	.	0.36	-0.20	*	.	F	1.25	0.49
Tyr	245	.	.	.	.	T	T	.	0.82	0.20	.	*	F	1.11	1.03
Gly	246	.	.	.	.	T	T	.	0.82	-0.06	*	*	F	2.02	1.03
Ala	247	.	.	.	.	T	.	.	0.97	-0.49	.	*	F	2.13	1.03
Gln	248	.	.	B	.	.	T	.	1.31	0.09	.	*	F	1.49	0.35
Gly	249	.	.	.	.	T	T	.	1.10	-0.67	.	*	F	3.10	0.59
Asp	250	.	.	.	.	T	T	.	1.34	-0.67	.	*	F	2.79	0.91
Cys	251	.	.	.	.	.	T	C	1.10	-1.17	.	*	F	2.28	0.91
Asp	252	.	.	.	.	.	.	C	1.48	-1.07	.	*	F	1.77	0.93
Pro	253	.	.	.	.	.	.	C	0.88	-1.07	.	*	F	1.46	0.86
Glu	254	.	A	.	.	.	.	C	0.91	-0.46	.	*	F	0.80	1.58
Ala	255	A	A	.	.	.	.	.	0.91	-0.54	.	*	F	0.90	1.37
Pro	256	A	A	.	.	.	.	.	1.23	-0.54	.	.	F	0.90	1.53
Met	257	A	A	.	.	.	.	.	0.92	-0.54	*	.	F	0.75	0.88
Thr	258	A	A	.	.	.	.	.	1.24	-0.06	*	.	F	0.60	1.25
Glu	259	A	A	.	.	.	.	.	0.58	-0.56	*	.	F	0.90	1.59
Gly	260	.	.	.	.	T	T	.	0.50	-0.41	*	.	F	1.25	0.86
Thr	261	A	.	.	.	.	T	.	0.82	-0.46	*	.	F	0.85	0.32
Arg	262	A	.	.	.	.	T	.	1.42	-0.94	*	.	F	1.15	0.36
Cys	263	A	.	.	.	.	T	.	1.73	-0.54	*	.	.	1.00	0.63
Cys	264	A	A	.	.	.	.	.	1.13	-0.97	*	.	.	0.60	0.76
Arg	265	A	A	.	.	.	.	.	1.23	-0.84	*	.	.	0.60	0.38

Table II (continued)

Res	Position	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Gln	266		A	B					0.66	-0.09	*	*	F	0.60	1.12
Glu	267		A	B					0.54	0.03		*		-0.15	1.46
Met	268		A	B					0.40	-0.54		*		0.75	1.25
Tyr	269		A	B					1.07	0.14		*		-0.30	0.59
Ile	270	A	A						0.61	0.14		*		-0.30	0.59
Asp	271	A	A						0.01	0.57		*		-0.60	0.59
Leu	272	A	A						0.06	0.57		*		-0.60	0.38
Gln	273	A	A						0.37	-0.19		*		0.45	1.07
Gly	274	A	A						0.02	0.04				-0.30	0.67
Met	275	A	A						0.91	0.54	*	*		-0.60	0.83
Lys	276	A	A						0.91	-0.14	*			0.30	0.83
Trp	277	A	A						1.43	-0.14	*			0.45	1.34
Ala	278	A	A						0.58	0.34	*			-0.15	1.43
Glu	279	A	A						0.11	0.37	*			-0.30	0.53
Asn	280	A	A						0.71	1.06	*	*		-0.60	0.42
Trp	281		A	B					0.46	0.14	*			-0.30	0.71
Val	282		A					C	0.53	0.07				-0.10	0.64
Leu	283		A					C	0.78	0.50				-0.40	0.61
Glu	284		A					C	0.08	0.53			F	-0.25	0.57
Pro	285						T	C	-0.73	0.40			F	0.45	0.67
Pro	286					T	T		-1.03	0.44			F	0.35	0.67
Gly	287					T	T		-0.42	0.26				0.50	0.39
Phe	288	A					T		0.39	1.01				-0.20	0.40
Leu	289	A	A						-0.28	0.59				-0.60	0.44
Ala	290	A	A	B					-0.92	0.73				-0.60	0.24
Tyr	291		A	B					-1.06	0.94				-0.60	0.21
Glu	292		A	B					-1.02	0.59				-0.60	0.25
Cys	293		A			T			-0.99	0.39		*		0.10	0.35
Val	294					T			-0.07	0.46	*			0.00	0.12
Gly	295					T	T		0.52	-0.30				1.10	0.14
Thr	296					T	T		0.56	0.10			F	0.95	0.44
Cys	297					T	T		0.34	-0.04	*		F	1.85	0.92
Arg	298					T	T		1.01	-0.26	*		F	2.30	1.44
Gln	299							C	1.28	-0.69	*		F	2.50	1.73
Pro	300						T	C	0.81	-0.67	*		F	3.00	3.25
Pro	301						T	C	0.53	-0.56	*		F	2.70	1.37
Glu	302	A					T		0.50	-0.06		*	F	1.75	0.80
Ala	303	A					T		0.43	0.33		*		0.70	0.45
Leu	304	A	A						0.14	-0.10				0.60	0.58
Ala	305	A	A						0.14	0.39		*		-0.30	0.35
Phe	306	A	A						-0.34	0.81				-0.60	0.54
Lys	307	A	A						-1.16	1.10		*		-0.60	0.56
Trp	308		A	B					-0.91	1.10		*		-0.60	0.46
Pro	309		A					C	-0.31	1.03	*	*		-0.40	0.53
Phe	310					T			0.39	0.67	*			0.00	0.41
Leu	311							C	1.09	0.67	*	*		-0.20	0.76
Gly	312						T	C	0.38	0.16	*	*	F	0.45	0.85
Pro	313					T	T		-0.22	0.30			F	0.65	0.53
Arg	314					T	T		-0.60	0.20			F	0.65	0.45
Gln	315					T	T		-0.20	0.01				0.50	0.46
Cys	316			B	B				0.61	-0.03				0.30	0.40
Ile	317			B	B				0.64	-0.46				0.64	0.35
Ala	318			B	B				0.86	0.03				0.38	0.29
Ser	319			B	B				0.44	-0.37	*		F	1.47	0.91
Glu	320			B			T		-0.37	-0.56	*		F	2.66	1.74
Thr	321					T	T		0.09	-0.56			F	3.40	1.42
Asp	322					T	T		0.38	-0.63			F	3.06	1.64
Ser	323	A					T		0.08	-0.40			F	1.87	0.93
Leu	324	A			B				-0.48	0.29				0.38	0.45
Pro	325	A			B				-0.78	0.44				-0.26	0.20
Met	326	A			B				-1.36	0.83		*		-0.60	0.20
Ile	327			B	B				-1.31	1.13				-0.60	0.17
Val	328			B	B				-1.01	0.44		*		-0.60	0.22
Ser	329			B					-0.54	0.01		*		0.24	0.39
Ile	330			B					-0.68	-0.17	*	*	F	1.33	0.55
Lys	331			B			T		0.03	-0.43	*	*	F	1.87	0.73

Table II (continued)

Res	Position	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Glu	332	.	.	.	.	T	T	.	0.61	-1.07	.	*	F	3.06	1.07
Gly	333	.	.	.	.	T	T	.	1.58	-0.97	.	*	F	3.40	2.20
Gly	334	.	.	.	.	T	T	.	1.67	-1.66	*	*	F	3.06	2.15
Arg	335	.	.	.	.	T	.	.	2.56	-1.23	*	*	F	2.52	1.92
Thr	336	.	.	.	.	.	.	C	1.66	-0.83	*	*	F	1.98	3.36
Arg	337	.	.	B	B	.	.	.	0.80	-0.61	*	*	F	1.24	2.52
Pro	338	.	.	B	B	.	.	.	0.84	-0.40	.	*	F	0.45	0.96
Gln	339	.	.	B	B	.	.	.	0.38	-0.01	.	*	.	0.30	0.89
Val	340	.	.	B	B	.	.	.	0.06	0.19	.	*	.	-0.30	0.37
Val	341	.	.	B	B	.	.	.	0.37	0.61	.	.	.	-0.60	0.37
Ser	342	.	.	B	.	.	.	.	-0.34	0.59	.	*	.	-0.40	0.35
Leu	343	.	.	B	.	.	T	.	-0.02	0.80	.	*	.	-0.20	0.46
Pro	344	.	.	B	.	.	T	.	-0.88	0.16	.	*	.	0.25	1.22
Asn	345	.	.	.	.	T	T	.	-0.02	0.16	.	*	.	0.50	0.68
Met	346	A	.	.	.	.	T	.	0.88	0.17	.	.	.	0.25	1.42
Arg	347	A	.	.	.	.	.	.	0.51	-0.51	.	*	.	0.95	1.84
Val	348	.	.	B	.	.	.	.	1.02	-0.37	.	*	.	0.50	0.61
Gln	349	.	.	B	.	.	T	.	0.57	-0.39	.	*	.	0.70	0.83
Lys	350	.	.	B	.	.	T	.	-0.02	-0.43	.	*	.	0.70	0.23
Cys	351	.	.	B	.	.	T	.	0.28	0.07	.	*	.	0.10	0.31
Ser	352	.	.	B	.	.	T	.	0.17	-0.19	.	*	.	0.70	0.24
Cys	353	.	.	B	.	.	.	.	0.68	-0.59	.	.	.	0.80	0.20
Ala	354	.	.	B	.	.	T	.	0.09	-0.16	.	.	.	0.70	0.37
Ser	355	.	.	.	.	T	T	.	-0.77	-0.23	.	.	.	1.10	0.28
Asp	356	.	.	.	.	T	T	.	-0.96	0.07	.	.	.	0.50	0.43
Gly	357	.	.	.	.	T	T	.	-0.87	0.14	.	.	.	0.50	0.31
Ala	358	.	.	B	.	.	.	.	-0.09	0.07	*	.	.	0.06	0.36
Leu	359	.	.	B	.	.	.	.	0.61	-0.31	*	.	.	0.82	0.42
Val	360	.	.	B	.	.	.	.	0.10	-0.31	*	.	.	0.98	0.84
Pro	361	.	.	B	.	.	.	.	0.10	-0.06	*	.	F	1.29	0.69
Arg	362	.	.	B	.	.	.	.	0.23	-0.16	*	.	F	1.60	1.44
Arg	363	.	.	B	.	.	.	.	0.43	-0.41	*	.	F	1.44	3.00
Leu	364	.	.	B	.	.	.	.	0.86	-0.63	*	.	.	1.43	2.48
Gln	365	.	.	B	.	.	.	.	1.32	-0.63	*	.	.	1.27	1.62
Pro	366	.	.	B	.	.	.	.	1.14	-0.20	*	.	.	0.81	1.06

Among highly preferred fragments in this regard are those that comprise regions of Human Nodal or Human Lefty that combine several structural features, such as, two, three, four, five or more of the features set out above.

In another embodiment, the invention provides isolated nucleic acid molecules comprising polynucleotides which hybridize under stringent hybridization conditions to a portion of the polynucleotide in a nucleic acid molecule of the inventions described above, for instance, the cDNA clones contained in ATCC Deposit Nos. 209092, 209135, and 209091 and/or a polynucleotide fragment described above. By "stringent hybridization conditions" is intended overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (750 mM NaCl, 75 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 μg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

Further specific embodiments are directed to polynucleotides corresponding to nucleotides 1-125, 1-90, 1-60, 1-30, 30-125, 30-90, 30-60, 60-125, 60-90, 90-125, 310-930, 350-930, 400-930, 450-930, 500-930, 550-930, 600-930, 650-930, 700-930, 750-930, 800-930, 850-930, 900-930, 310-900, 350-900, 400-900, 450-900, 500-900, 550-900, 600-900, 650-900, 700-900, 750-900, 800-900, 850-900, 310-850, 350-850, 400-850, 450-850, 500-850, 550-850, 600-850, 650-850, 700-850, 750-850, 800-850, 310-800, 350-800, 400-800, 450-800, 500-800, 550-800, 600-800, 650-800, 700-800, 750-800, 310-750, 350-750, 400-750, 450-750, 500-750, 550-750, 600-750, 650-750, 700-750, 310-700, 350-700, 400-700, 450-700, 500-700, 550-700, 600-700, 650-700, 310-650, 350-650, 400-650, 450-650, 500-650, 550-650, 600-650, 310-600, 350-600, 400-600, 450-600, 500-600, 550-600, 310-500, 350-500, 400-500, 450-500, 310-450, 350-450, 400-450, 310-400, 350,-400, 310-350, 1050-1596, 1100-1596, 1150-1596, 1200-1596, 1250-1596, 1300-1596,

1350-1596, 1400-1596, 1450-1596, 1500-1596, 1550-1596, 1050-1550, 1100-1550, 1150-1550, 1200-1550, 1250-1550, 1300-1550, 1350-1550, 1400-1550, 1450-1550, 1500-1550, 1050-1500, 1100-1500, 1150-1500, 1200-1500, 1250-1500, 1300-1500, 1350-1500, 1400-1500, 1450-1500, 1050-1450, 1100-1450, 1150-1450, 1200-1450, 1250-1450, 1300-1450, 1350-1450, 1400-1450, 1050-1400, 1100-1400, 1150-1400, 1200-1400, 1250-1400, 1300-1400, 1350-1400, 1050-1350, 1100-1350, 1150-1350, 1200-1350, 1250-1350, 1300-1350, 1050-1300, 1100-1300, 1150-1300, 1200-1300, 1250-1300, 1050-1250, 1100-1250, 1150-1250, 1200-1250, 1050-1200, 1100-1200, 1150-1200, 1050-1150, 1100-1150, and 1050-1100 of SEQ ID NO:3.

By a polynucleotide which hybridizes to a "portion" of a polynucleotide is intended a polynucleotide (either DNA or RNA) hybridizing to at least about 15 nucleotides (nt), and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably about 30-70 (e.g., 50) nt of the reference polynucleotide. These are useful as diagnostic probes and primers as discussed above and in more detail below.

By a portion of a polynucleotide of "at least 20 nt in length," for example, is intended 20 or more contiguous nucleotides from the nucleotide sequence of the reference polynucleotides (e.g., the deposited cDNAs or the nucleotide sequences as shown in Figures 1A and B and 2A and B (SEQ ID NO:1 and SEQ ID NO:3, respectively)). Of course, a polynucleotide which hybridizes only to a poly A sequence (such as the 3' terminal poly(A) tract of the Nodal and Lefty cDNAs shown in Figures 1A and B and 2A and B (SEQ ID NO:1 and SEQ ID NO:3, respectively)), or to a complementary stretch of T (or U) residues, would not be included in a polynucleotide of the invention used to hybridize to a portion of a nucleic acid of the invention, since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof

(e.g., practically any double-stranded cDNA clone generated using oligo dT as a primer).

In preferred embodiments, polynucleotides which hybridize to the reference polynucleotides disclosed herein encode polypeptides which either retain substantially the same biological function or activity as the mature form or TGF-β-like active form of the Nodal polypeptide encoded by the polynucleotide sequences depicted in Figures 1A and 1B (SEQ ID NO:1) and/or substantially the same biological function or activity as the mature form or TGF-β-like active forms of the Lefty polypeptide encoded by the polynucleotide sequences depicted in Figures 2A and 2B (SEQ ID NO:1) depicted in Figures 2A and 2B (SEQ ID NO:3), or the cDNAs contained in the deposit (HTLFA20, HNGEF08, and HUKEJ46).

Alternative embodiments are directed to polynucleotides which hybridize to the reference polynucleotide (i.e., a polynucleotide sequence disclosed herein), but do not retain biological activity. While these polynucleotides do not retain biological activity, they have uses, such as, for example, as probes for the polynucleotides of SEQ ID NO:1 or SEQ ID NO:3, for recovery of the polynucleotides, as diagnostic probes, and as PCR primers.

As indicated, nucleic acid molecules of the present invention which encode a Lefty polypeptide may include, but are not limited to those encoding the amino acid sequence of the mature form of the polypeptide, by itself; and the coding sequence for the mature form of the polypeptide and additional sequences, such as those encoding the about 18 amino acid leader or secretory sequence, such as a pre-, or pro- or prepro- protein sequence; the coding sequence of the mature polypeptide, with or without the aforementioned additional coding sequences.

As indicated, nucleic acid molecules of the present invention which encode a Nodal polypeptide may include, but are not limited to, those encoding the amino acid sequence of the complete polypeptide, by itself; and the coding

sequence for the complete polypeptide and additional sequences, such as those encoding an added secretory leader sequence, such as a pre-, or pro- or prepro-protein sequence.

Also encoded by nucleic acids of the invention are the above protein sequences together with additional, non-coding sequences, including for example, but not limited to introns and non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences that play a role in transcription, mRNA processing, including splicing and polyadenylation signals, for example - ribosome binding and stability of mRNA; an additional coding sequence which codes for additional amino acids, such as those which provide additional functionalities.

Thus, the sequences encoding the polypeptides may be fused to a marker sequence, such as a sequence encoding a peptide which facilitates purification of the fused polypeptide. In certain preferred embodiments of the invention, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described by Gentz and colleagues (*Proc. Natl. Acad. Sci. USA* **86**:821-824 (1989)), for instance, hexa-histidine provides for convenient purification of the fusion protein. The "HA" tag is another peptide useful for purification which corresponds to an epitope derived from the influenza hemagglutinin protein, which has been described by Wilson and coworkers (*Cell* **37**:767 (1984)). As discussed below, other such fusion proteins include the Nodal and Lefty fused to Fc at the N- or C-terminus.

The present invention further relates to variants of the nucleic acid molecules of the present invention, which encode portions, analogs or derivatives of the Nodal and Lefty proteins. Variants may occur naturally, such as a natural allelic variant. By an "allelic variant" is intended one of several alternate forms of a gene occupying a given locus on a chromosome of an organism (*Genes II*, Lewin,

B., ed., John Wiley & Sons, New York (1985)). Non-naturally occurring variants may be produced using art-known mutagenesis techniques.

Such variants include those produced by nucleotide substitutions, deletions or additions. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding regions, non-coding regions, or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the Nodal and Lefty proteins or portions thereof. Also especially preferred in this regard are conservative substitutions.

Most highly preferred are nucleic acid molecules encoding the mature form of the protein having the amino acid sequence shown in SEQ ID NO:4 or the mature Lefty amino acid sequence encoded by the deposited cDNA clone.

Most highly preferred are nucleic acid molecules encoding the active domain of the proteins having the amino acid sequence shown in SEQ ID NO:2 or SEQ ID NO:4 or the active domains of the Nodal and Lefty amino acid sequences encoded by the deposited cDNA clones. By "active domain", is meant the C-terminal region of a Nodal or Lefty polypeptide, or fragment thereof, which has been processed either *in vitro* or *in vivo* such that the C-terminal region has been cleaved from the remainder of the molecule just C-terminal to one or more of the TGF-β cleavage consensus sites as indicated in Figures 1A and 1B and 2A and 2B.

Further embodiments include an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 90% identical, and more preferably at least 95%, 96%, 97%, 98% or 99% identical to a polynucleotide selected from the group consisting of: (a) a nucleotide sequence encoding the Nodal polypeptide having the complete amino acid sequence in SEQ ID NO:2 (i.e., positions 1 to 283 of SEQ ID NO:2); (b) a nucleotide sequence encoding the

predicted active Nodal polypeptide having the amino acid sequence at positions 173 to 283 of SEQ ID NO:2; (c) a nucleotide sequence encoding the Nodal polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 and/or 209135; (d) a nucleotide sequence encoding the active domain of the Nodal polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 and/or 209135; (e) a nucleotide sequence encoding the Lefty polypeptide having the complete amino acid sequence in SEQ ID NO:4 (i.e., positions -18 to 348 of SEQ ID NO:4); (f) a nucleotide sequence encoding the Lefty polypeptide having the complete amino acid sequence in SEQ ID NO:4 excepting the N-terminal methionine (i.e., positions -17 to 348 of SEQ ID NO:4); (g) a nucleotide sequence encoding the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 60 to 348 of SEQ ID NO:4; (h) a nucleotide sequence encoding the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 118 to 348 of SEQ ID NO:4; (i) a nucleotide sequence encoding the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 125 to 348 of SEQ ID NO:4; (j) a nucleotide sequence encoding the Lefty polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091; (k) a nucleotide sequence encoding the Lefty polypeptide having the complete amino acid sequence excepting the N-terminal methionine encoded by the cDNA clone contained in ATCC Deposit No. 209091; (l) a nucleotide sequence encoding the active domain of the Lefty polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091; and (m) a nucleotide sequence complementary to any of the nucleotide sequences in (a) through (l) above.

Further embodiments of the invention include isolated nucleic acid molecules that comprise a polynucleotide having a nucleotide sequence at least

90% identical, and more preferably at least 95%, 96%, 97%, 98% or 99% identical, to any of the nucleotide sequences in (a) through (m) above, or a polynucleotide which hybridizes under stringent hybridization conditions to a polynucleotide in (a) through (m) above. This polynucleotide which hybridizes does not hybridize under stringent hybridization conditions to a polynucleotide having a nucleotide sequence consisting of only A residues or of only T residues. An additional nucleic acid embodiment of the invention relates to an isolated nucleic acid molecule comprising a polynucleotide which encodes the amino acid sequence of an epitope-bearing portion of a Nodal and Lefty polypeptide having an amino acid sequence in (a) through (l) above. A further nucleic acid embodiment of the invention relates to an isolated nucleic acid molecule comprising a polynucleotide which encodes the amino acid sequence of a Human Nodal or Human Lefty polypeptide having an amino acid sequence which contains at least one conservative amino acid substitution, but not more than 50 conservative amino acid substitutions, even more preferably, not more than 40 conservative amino acid substitutions, still more preferably not more than 30 conservative amino acid substitutions, and still even more preferably not more than 20 conservative amino acid substitutions. Of course, in order of ever-increasing preference, it is highly preferable for a polynucleotide which encodes the amino acid sequence of a Human Nodal or Human Lefty polypeptide to have an amino acid sequence which contains not more than 7-10, 5-10, 3-7, 3-5, 2-5, 1-5, 1-3, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 conservative amino acid substitutions.

By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence encoding a Nodal or Lefty polypeptide is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference

nucleotide sequences encoding the Nodal and Lefty polypeptides. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular nucleic acid molecule is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the nucleotide sequences shown in Figures 1A and B and 2A and B or to the nucleotides sequence of the deposited cDNA clones can be determined conventionally using known computer programs such as the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711). Bestfit uses the local homology algorithm of Smith and Waterman to find the best segment of homology between two sequences (*Advances in Applied Mathematics* **2**:482-489 (1981)). When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such that the percentage of identity is calculated over the full length of the reference nucleotide sequence and that gaps in homology of up to 5% of the total number of nucleotides in the reference sequence are allowed. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the

algorithm of Brutlag and colleagues (*Comp. App. Biosci.* **6**:237-245 (1990)). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a

matched/alignment of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

The present application is directed to nucleic acid molecules at least 90%, 95%, 96%, 97%, 98% or 99% identical to the nucleic acid sequences shown in Figures 1A and B and 2A and B (SEQ ID NO:1 and SEQ ID NO:3, respectively) or to the nucleic acid sequences of the deposited cDNAs, irrespective of whether they encode a polypeptide having Nodal or Lefty activity. This is because even where a particular nucleic acid molecule does not encode a polypeptide having Nodal or Lefty activity, one of skill in the art would still know how to use the nucleic acid molecule, for instance, as a hybridization probe or a polymerase chain reaction (PCR) primer. Uses of the nucleic acid molecules of the present invention that do not encode a polypeptide having Nodal or Lefty activity include, *inter alia*, (1) isolating the Nodal or Lefty genes or allelic variants thereof in a cDNA library; (2) *in situ* hybridization (e.g., "FISH") to metaphase chromosomal spreads to provide precise chromosomal location of the Nodal or Lefty genes, as described by Verma and colleagues (*Human Chromosomes: A Manual of Basic Techniques*, Pergamon Press, New York (1988)); and Northern Blot analysis for detecting Nodal or Lefty mRNA expression in specific tissues.

Preferred, however, are nucleic acid molecules having sequences at least 90%, 95%, 96%, 97%, 98% or 99% identical to the nucleic acid sequences shown in Figures 1A and B and 2A and B (SEQ ID NO:1 and SEQ ID NO:3, respectively) or to the nucleic acid sequences of the deposited cDNAs or to fragments of these polynucleotides as described herein, which do, in fact, encode polypeptides having Nodal or Lefty activity. By "a polypeptide having Nodal or Lefty activity" is intended polypeptides exhibiting activity similar, but not necessarily identical, to an activity of the active forms of Nodal or Lefty proteins of the invention, as measured in a particular biological assay. For example, the Nodal and Lefty proteins of the present invention are involved in the regulation of cell growth and differentiation. Other TGF-β-like molecules have the capacity to stimulate the proliferation of human endothelial cells in the presence of the comitogen concanavalin A (conA). Such an activity may be easily assayed by directly examining the effects of Nodal or Lefty or any muteins thereof on the proliferation of human endothelial cells as follows. Endothelial cells are obtained and cultured in 96 well flat-bottomed culture dishes (Costar, Cambridge, MA) in RPMI 1640 medium supplemented with 10% heat-inactivated fetal bovine serum (HyClone Labs, Logan, UT), 1% L-glutamine, 100 U/mL penicillin, 100 μg/mL streptomycin, 0.1% gentamicin (Life Technologies, Inc., Rockville, MD) in the presence of 2 μg/mL conA (Calbiochem, La Jolla, CA). ConA and the polypeptide to be analyzed are added to a final volume of medium of 0.2 mL. After 60 h at 37°C, cultures are pulsed with 1 μCi of [^{3}H]-thymidine (5 Ci/mmol; 1 Ci=37 BGq; NEN) for 12-18 h and harvested onto glass fiber filters (PhD; Cambridge Technology, Watertown, MA). Mean [^{3}H]-thymidine incorporation (CPM) of triplicate cultures is determined using a liquid scintillation counter (Beckman Instruments, Irvine, CA). Significant [^{3}H]-thymidine incorporation indicates stimulation of endothelial cell proliferation. Such activity is useful for

determining the potential for inducing or repressing the capacity for cellular growth and proliferation that Nodal or Lefty or a mutein thereof may possess.

Nodal and Lefty proteins regulate cellular proliferation and differentiation in a dose-dependent manner in the above-described assays. Although the compositions of the invention need not regulate cellular proliferation and differentiation in a dose-dependent manner, it is preferred that "a polypeptide having Nodal or Lefty activity" includes polypeptides that also exhibit any of the same cellular proliferation and differentiation regulatory activities in the above-described assays in a dose-dependent manner. Although the degree of dose-dependent activity need not be identical to that of the Nodal or Lefty proteins, preferably, "a polypeptide having Nodal or Lefty protein activity" will exhibit substantially similar dose-dependence in a given activity as compared to the Nodal or Lefty proteins (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity relative to the reference Nodal and Lefty proteins).

Further analysis of the ability of polypeptides of the invention to regulate cellular growth or differentiation of a particular cell type may be ascertained through the use of an *in vitro* colony forming assay to measure the extent of inhibition of myeloid progenitor cells (Youn, *et al., J. Immunol.* **155**:2661-2667 (1995)). Briefly, this assay involves collecting human or mouse bone marrow cells and plating the same on agar, adding one or more growth factors and either (1) transfected host cell-supernatant containing Nodal or Lefty protein (or a candidate polypeptide) or (2) nontransfected host cell-supernatant control, and measuring the effect on colony formation by murine and human CFU-granulocyte-macrophages (CFU-GM), by human burst-forming unit-erythroid (BFU-E), or by human CFU granulocyte-erythroid-macrophage-megakaryocyte (CFU-GEMM).

Like other TGF-β-related molecules, Nodal and Lefty may exhibit an activity on leukocytes including, for example, monocytes, lymphocytes and neutrophils. For this reason, Nodal and Lefty are active in directing the proliferation and differentiation of these cell types. Such activity is useful, for example, for immune enhancement or suppression, myeloprotection, stem cell mobilization, acute and chronic inflammatory control and treatment of leukemia. Assays for measuring such activity are well known in the art (Peters, *et al., Immun. Today* **17:**273 (1996); Young, *et al., J. Exp. Med.* **182:**1111 (1995); Caux, *et al., Nature* **390:**258 (1992); and Santiago-Schwarz, *et al., Adv. Exp. Med. Biol.* **378:**7 (1995).

Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of the nucleic acid molecules having a sequence at least 90%, 95%, 96%, 97%, 98%, or 99% identical to the nucleic acid sequence of the deposited cDNA or the nucleic acid sequences shown in Figures 1A and B and 2A and B (SEQ ID NO:1 and SEQ ID NO:3, respectively), or fragments thereof, will encode polypeptides "having Nodal or Lefty protein activity." In fact, since degenerate variants of these nucleotide sequences all encode the same polypeptides, this will be clear to the skilled artisan even without performing the above described comparison assay. It will be further recognized in the art that, for such nucleic acid molecules that are not degenerate variants, a reasonable number will also encode a polypeptide having Nodal or Lefty activity. This is because the skilled artisan is fully aware of amino acid substitutions that are either less likely or not likely to significantly effect protein function (e.g., replacing one aliphatic amino acid with a second aliphatic amino acid), as further described below.

Polynucleotide Assays

The invention also encompasses the use of Nodal and Lefty polynucleotides to detect complementary polynucleotides, such as, for example, as a diagnostic reagent for detecting diseases or susceptibility to diseases related to the presence of mutated Nodal and Lefty. Such diseases are related to an under-expression of Nodal and Lefty, such as, for example, abnormal cellular proliferation such as tumors and cancers.

Individuals carrying mutations in the human Nodal or Lefty genes may be detected at the DNA level by a variety of techniques. Nucleic acids for diagnosis may be obtained from a patient's cells, such as from blood, urine, saliva, tissue biopsy and autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR (Saiki *et al., Nature* **324:**163-166 (1986)) prior to analysis. RNA or cDNA may also be used for the same purpose. As an example, PCR primers complementary to the nucleic acid encoding Nodal or Lefty can be used to identify and analyze Nodal or Lefty mutations. For example, deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to radiolabeled Nodal or Lefty RNA or alternatively, radiolabeled Nodal or Lefty antisense DNA sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase A digestion or by differences in melting temperatures.

Genetic testing based on DNA sequence differences may be achieved by detection of alteration in electrophoretic mobility of DNA fragments in gels with or without denaturing agents. Small sequence deletions and insertions can be visualized by high resolution gel electrophoresis. DNA fragments of different sequences may be distinguished on denaturing formamide gradient gels in which the mobilities of different DNA fragments are retarded in the gel at different positions according to their specific melting or partial melting temperatures (see, e.g., Myers *et al.*, *Science* **230:**1242 (1985)).

Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (e.g., Cotton *et al.*, *Proc. Natl. Acad. Sci., USA*, **85**:4397-4401 (1985)).

Thus, the detection of a specific DNA sequence may be achieved by methods such as hybridization, RNase protection, chemical cleavage, direct DNA sequencing or the use of restriction enzymes, (e.g., Restriction Fragment Length Polymorphisms (RFLP)) and Southern blotting of genomic DNA.

In addition to more conventional gel-electrophoresis and DNA sequencing, mutations can also be detected by *in situ* analysis.

Vectors and Host Cells

While the Lefty and Nodal polypeptides (including fragments, variants derivatives, and analogs) of the invention can be chemically synthesized (e.g., see Creighton, 1983, Proteins: Structures and Molecular Principles, W.H. Freeman & Co., N.Y.), Lefty and Nodal polypeptides may advantageously be produced by recombinant DNA technology using techniques well known in the art for expressing gene sequences and/or nucleic acid coding sequences. Such methods can be used to construct expression vectors containing the polynucleotides of the invention and appropriate transcriptional and translational control signals. These methods include, for example, in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. See, for example, the techniques described in Sambrook et al., 1989, supra; Ausubel et al., 1989, supra; Caruthers et al., 1980, Nuc. Acids Res. Symp. Ser. 7:215-233; Crea and Horn, 1980, Nuc. Acids Res. 9(10):2331; Matteucci and Caruthers, 1980, Tetrahedron Letters 21:719; and Chow and Kempe, 1981, Nuc. Acids Res. 9(12):2807-2817. Alternatively, RNA capable of Lefty or Nodal sequences may be chemically synthesized using, for example, synthesizers. See, for example, the

techniques described in "Oligonucleotide Synthesis", 1984, Gait, M.J. ed., IRL Press, Oxford, which is incorporated by reference herein in its entirety.

Thus, in one embodiment, the present invention relates to vectors which include the isolated DNA molecules (i.e., polynucleotides) of the present invention, host cells which are genetically engineered with the recombinant vectors, and the production of Nodal or Lefty polypeptides or fragments thereof by recombinant techniques using these host cells or host cells that have otherwise been genetically engineered using techniques known in art to express a polypeptide of the invention. The vector may be, for example, a phage, plasmid, viral or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged *in vitro* using an appropriate packaging cell line and then transduced into host cells.

In one embodiment, the polynucleotide of the invention is operatively associated with an appropriate heterologous regulatory element (e.g., a promoter or enhancer or both), such as the phage lambda PL promoter, the *E. coli lac*, *trp*, *phoA* and *tac* promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan.

In embodiments in which vectors contain expression constructs, these constructs will further contain sites for transcription initiation, termination and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA,

UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, 293 and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Vectors preferred for use in bacteria include pHE4-5, pQE70, pQE60 and pQE-9 (QIAGEN, Inc., *supra*); pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1, and pSG (Stratagene); and pSVK3, pBPV, pMSG and pSVL (Pharmacia). Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection or other methods. Such methods are described in many standard laboratory manuals (for example, Davis, *et al., Basic Methods In Molecular Biology* (1986)).

In addition to encompassing host cells containing the vector constructs discussed herein, the invention also encompasses primary, secondary, and immortalized host cells of vertebrate origin, particularly those of mammalian origin, that have been engineered to delete or replace endogenous genetic material

(e.g., Human Nodal or Human Lefty coding sequence), and/or to include genetic material (e.g. heterologous polynucleotide sequences) that is operably associated with Human Nodal or Human Lefty polynucleotides of the invention, and which activates, alters, and/or amplifies endogenous Human Nodal or Human Lefty polynucleotides. For example, techniques known in the art may be used to operably associate heterologous control regions (e.g. promoter and/or enhancer) and endogenous Human Nodal or Human Lefty polynucleotide sequences via homologous recombination (see, e.g. U.S. Patent No. 5,641,670, issued June 24, 1997; International Publication No. WO 96/29411, published September 26, 1996; International Publication No. WO 94/12650, published August 4, 1994; Koller et al., Proc. Natl. Acad. Sci. USA 86:8932-8935 (1989); and Zijlstra, et al., Nature 342:435-438 (1989), the disclosures of each of which are hereby incorporated by reference in their entireties).

The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals, but also additional heterologous functional regions. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence in the host cell, during purification, or during subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to polypeptides to engender secretion or excretion, to improve stability and to facilitate purification, among others, are familiar and routine techniques in the art. A preferred fusion protein comprises a heterologous region from immunoglobulin that is useful to stabilize and purify proteins. For example, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part

in a fusion protein is thoroughly advantageous for use in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232 262). On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected and purified in the advantageous manner described. This is the case when Fc portion proves to be a hindrance to use in therapy and diagnosis, for example when the fusion protein is to be used as antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5 (Bennett, D., *et al., J. Molecular Recognition* **8**:52-58 (1995); Johanson, K., *et al., J. Biol. Chem.* **270**:9459-9471 (1995)).

The Nodal and Lefty proteins can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification. Polypeptides of the present invention include: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon

generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

Included within the scope of the invention are Lefty and Nodal polypeptides (including fragments, variants, derivatives and analogs) which are differentially modified during or after translation, e.g., by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, etc. Any of numerous chemical modifications may be carried out by known techniques, including, but not limited to, specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH4; acetylation, formylation, oxidation, reduction; metabolic synthesis in the presence of tunicamycin; etc. In a specific embodiment, the compositions of the invention are conjugated to other molecules to increase their water-solubility (e.g., polyethylene glycol), half-life, or ability to bind targeted tissue (e.g., bisphosphonates and fluorochromes to target the proteins to bony sites).

Polypeptides and Fragments

The invention further provides isolated Nodal and Lefty polypeptides having the amino acid sequences encoded by the deposited cDNAs, or the amino acid sequences in SEQ ID NO:2 and SEQ ID NO:4, respectively, or a peptide or polypeptide comprising a fragment (i.e., a portion) of the above polypeptides.

The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to a point within the range of near complete (e.g., >90% pure) to complete (e.g., >99%

pure) homogeneity. The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Also intended as an "isolated polypeptide" are polypeptides that have been purified partially or substantially from a recombinant host cell. For example, a recombinantly produced version of a Nodal or Lefty polypeptide can be substantially purified by the one-step method described by Smith and Johnson (*Gene* **67**:31-40 (1988)). Such polynucleotides could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment. Isolated polypeptides and polynucleotides according to the present invention also include such molecules produced naturally or synthetically. Polypeptides and polynucleotides of the invention also can be purified from natural or recombinant sources using anti-Nodal or anti-Lefty antibodies of the invention which may routinely be generated and utilized using methods known in the art.

To improve or alter the characteristics of Nodal and Lefty polypeptides, protein engineering may be employed. Recombinant DNA technology known to those skilled in the art can be used to create novel mutant proteins or muteins including single or multiple amino acid substitutions, deletions, additions or fusion proteins. Such modified polypeptides can show, e.g., enhanced activity or increased stability. In addition, they may be purified in higher yields and show better solubility than the corresponding natural polypeptide, at least under certain purification and storage conditions.

The present invention also encompasses fragments of the above-described Nodal and Lefty polypeptides. Polypeptide fragments of the present invention include polypeptides comprising an amino acid sequence contained in SEQ ID

NO:2, SEQ ID NO:4, encoded by the cDNA contained in the deposited clones (HTLFA20 and HNGEF08, (encoding Nodal) and HUKEJ46 (encoding Lefty)), or encoded by nucleic acids which hybridize (e.g., under stringent hybridization conditions) to the nucleotide sequence contained in the deposited clones, that shown in Figures 1A and 1B (SEQ ID NO:1) and/or Figures 2A and 2B (SEQ ID NO:3), or the complementary strand thereto.

Polypeptide fragments may be "free-standing" or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, included, for example, fragments that comprise or alternatively, consist of, from about amino acid residues, 1 to 20, 21 to 40, 41 to 60, 61 to 83, 84 to 100, 101 to 120, 121 to 140, 141 to 160, 161 to 180, 181 to 200, 201 to 220, 201 to 224, 210 to 231, 221 to 240, 241 to 260, 261 to 280, 261 to 283, 281 to 289, 281 to 300, 301 to 320, 321 to 340, 341 to 348, 341 to 360, and 341 to 366 of SEQ ID NO:2 and/or SEQ ID NO:4. Moreover, polypeptide fragments can be at least about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350 or 360 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (i.e. 5, 4, 3, 2 or 1) amino acids, at either extreme or at both extremes.

In other embodiments, the fragments or polypeptides of the invention (i.e., those described herein) are not larger than 325, 300, 250, 225, 200, 185, 175, 170, 165, 160, 155, 150, 145, 140, 135, 130, 125, 120, 115, 110, 105, 100, 90, 80, 75, 60, 50, 40, 30 or 25 amino acids residues in length.

Additional embodiments encompass polypeptide fragments comprising one or more functional regions of Nodal or Lefty polypeptides of the invention, such as, one or more Garnier-Robson alpha-regions, beta-regions, turn-regions, and coil-regions, Chou-Fasman alpha-regions, beta-regions, and coil-regions,

Kyte-Doolittle hydrophilic regions and hydrophobic regions, Eisenberg alpha- and beta-amphipathic regions, Karplus-Schulz flexible regions, Emini surface-forming regions and Jameson-Wolf regions of high antigenic index, or any combination thereof, as disclosed in Figures 5 and 6 and in Tables I and II and as described herein.

Further preferred embodiments encompass polypeptide fragments comprising, or alternatively consisting of, the TGF-β-like domain of Nodal (amino acid residues 174-283 of SEQ ID NO:2).

Additional preferred embodiments encompass polypeptide fragments comprising, or alternatively consisting of, the mature domain of Lefty (amino acid residues 1-348 of SEQ ID NO:4), the first predicted TGF-β-like domain of Lefty (amino acid residues 60-348 of SEQ ID NO:4), the second predicted TGF-β-like domain of Lefty (amino acid residues 118-348 of SEQ ID NO:4), and/or the third predicted TGF-β-like domain of Lefty (amino acid residues 125-348 of SEQ ID NO:4).

In specific embodiments, polypeptide fragments of the invention comprise, or alternatively, consist of, amino acid residues aspartic acid-1 to alanine-27, arginine-30 to glutamic acid-58, cysteine-64 to phenylalanine-82, glycine-85 to serine-110, and leucine-130 to leucine-283 of the Nodal sequence recited in SEQ ID NO:2. In additional specific embodiments, polypeptide fragments of the invention comprise, or alternatively, consist of, amino acid residues leucine-(-15) to serine-(-2), alanine-3 to leucine-19, valine-34 to histidine-51, arginine-54 to leucine-72, glutamic acid-75 to arginine-114, arginine-117 to proline-192, histidine-198 to proline-209, glycine-211 to leucine-286, tryptophan-290 to glutamic acid-302, and serine-305 to proline-348 of the Lefty amino acid sequence recited in SEQ ID NO:4. These domains are regions of high identity identified by comparison of the TNF family member polypeptides shown in Figures 3 and 4.

In additional specific embodiments, the polypeptides of the invention comprise, or alternatively consist of, amino acid residues 19 to 25, 84 to 104, 105-125, 126 to 150, 151 to 170, 171 to 200, 201-250, 251 to 270, 271 to 297, 329 to 339, and/or 340 363 of the Lefty amino acid sequence depicted in Figures 2A and 2B. Polynucleotides encoding these polypeptides are also encompassed by the invention, as are polynucleotides that hybridize to the complementary strand of these encoding polynucleotides under high stringency conditions (e.g., as described herein) and polypeptides encoded by these hybridizing polynucleotides.

The polypeptides of the present invention have uses which include, but are not limited to, a molecular weight marker on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art.

As described in detail below, the polypeptides of the present invention can also be used to raise polyclonal and monoclonal antibodies, which are useful in assays for detecting Nodal or Lefty protein expression as described below or as agonists and antagonists capable of enhancing or inhibiting Nodal or Lefty protein function. Further, such polypeptides can be used in the yeast two-hybrid system to "capture" Nodal or Lefty protein binding proteins which are also candidate agonists and antagonists according to the present invention. The yeast two hybrid system is described by Fields and Song (*Nature* **340:**245-246 (1989)).

In another embodiment, the invention provides peptides or polypeptides comprising epitope-bearing portions of a polypeptide of the invention. The epitope of this polypeptide portion is an immunogenic or antigenic epitope of a polypeptide of the invention. An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response when the whole protein is the immunogen. On the other hand, a region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope". The number of immunogenic epitopes of a protein generally is less than the number of antigenic

epitopes (see, for instance, Geysen, *et al., Proc. Natl. Acad. Sci. USA* **81**:3998-4002 (1983)).

As to the selection of peptides or polypeptides bearing an antigenic epitope (i.e., that contain a region of a protein molecule to which an antibody can bind), it is well known in that art that relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein (see, for instance, Sutcliffe, J. G., *et al., Science* **219**:660-666 (1983)). Peptides capable of eliciting protein-reactive sera are frequently represented in the primary sequence of a protein, can be characterized by a set of simple chemical rules, and are confined neither to immunodominant regions of intact proteins (i.e., immunogenic epitopes) nor to the amino or carboxyl terminals. Antigenic epitope-bearing peptides and polypeptides of the invention are therefore useful to raise antibodies, including monoclonal antibodies, that bind specifically to a polypeptide of the invention (see, for instance, Wilson, *et al., Cell* **37**:767-778 (1984)).

Antigenic epitope-bearing peptides and polypeptides of the invention preferably contain a sequence of at least seven, more preferably at least nine and most preferably between about 15 to about 30 amino acids contained within the amino acid sequence of a polypeptide of the invention. Non-limiting examples of antigenic polypeptides or peptides that can be used to generate Nodal-specific antibodies include: a polypeptide comprising amino acid residues from about Lys-54 to about Asp-62, from about Val-91 to about Leu-99, from about Lys-100 to about Gln-108, from about Cys-116 to about Pro-124, from about Gln-140 to about Leu-148, from about Trp-156 to about Ser-164, from about Arg-170, to about Gln-181, from about Cys-212 to about Phe-224, from about Tyr-239, to about Thr-247, from about Pro-251, to about Met-259, and from about Asp-263, to about His-271. Non-limiting examples of antigenic polypeptides or peptides that can be used to generate Lefty-specific antibodies

include: a polypeptide comprising amino acid residues from about Asp-71 to about Ser-79, from about Arg-106 to about Val-114, from about Leu-136 to about Arg-144, from about Asp-154 to about Asp-164, from about His-171 to about Asp-179, from about Gln-189 to about Leu-197, from about Pro-227 to about Glu-236, from about Gly-246 to about Glu-254, from about Pro-256 to about Gln-266, from about Cys-297 to about Ala-305, from about Ile-317 to about Pro-325, from about Ile-330 to about Val-340, and from about Val-348 to about Pro-366. These polypeptide fragments have been determined to bear antigenic epitopes of the Nodal and Lefty proteins by the analysis of the Jameson-Wolf antigenic index, as shown in Figures 5 and 6, and Tables I and II, above.

The epitope-bearing peptides and polypeptides of the invention may be produced by any conventional means (see, for example, Houghten, R. A., *et al.*, *Proc. Natl. Acad. Sci. USA 82*:5131-5135 (1985); and U.S. Patent No. 4,631,211 to Houghten, *et al.* (1986)).

Epitope-bearing peptides and polypeptides of the invention are used to induce antibodies according to methods well known in the art (see, for instance, Sutcliffe, *et al., supra*; Wilson, *et al., supra*; Chow, M., *et al., Proc. Natl. Acad. Sci. USA* **82**:910-914; and Bittle, F. J., *et al., J. Gen. Virol.* **66**:2347-2354 (1985)). Immunogenic epitope-bearing peptides of the invention, i.e., those parts of a protein that elicit an antibody response when the whole protein is the immunogen, are identified according to methods known in the art (see, for instance, Geysen, *et al., supra*). Further still, U.S. Patent No. 5,194,392, issued to Geysen, describes a general method of detecting or determining the sequence of monomers (amino acids or other compounds) which is a topological equivalent of the epitope (i.e., a "mimotope") which is complementary to a particular paratope (antigen binding site) of an antibody of interest. More generally, U.S. Patent No. 4,433,092, issued to Geysen, describes a method of detecting or determining a sequence of monomers which is a topographical equivalent of a ligand which is

complementary to the ligand binding site of a particular receptor of interest. Similarly, U.S. Patent No. 5,480,971, issued to Houghten and colleagues, on Peralkylated Oligopeptide Mixtures discloses linear C1-C7-alkyl peralkylated oligopeptides and sets and libraries of such peptides, as well as methods for using such oligopeptide sets and libraries for determining the sequence of a peralkylated oligopeptide that preferentially binds to an acceptor molecule of interest. Thus, non-peptide analogs of the epitope-bearing peptides of the invention also can be made routinely by these methods.

For many proteins, including the extracellular domain of a membrane associated protein or the mature form(s) of a secreted protein, it is known in the art that one or more amino acids may be deleted from the N-terminus or C-terminus without substantial loss of biological function. For instance, Ron and colleagues (*J. Biol. Chem.*, **268**:2984-2988 (1993)) reported modified KGF proteins that had heparin binding activity even if 3, 8, or 27 N-terminal amino acid residues were missing. In the present case, since the Nodal and Lefty proteins of the invention are members of the TGF-β polypeptide superfamily, deletions of N-terminal amino acids up to the N-terminal-most cysteine of the predicted active form of the proteins at positions 183 and 233 of SEQ ID NO:2 and SEQ ID NO:4, respectively, may retain some biological activity such as receptor binding or modulation of target cell activities. Polypeptides having further N-terminal deletions including the Cys-183 and Cys-233 residues in SEQ ID NO:2 and SEQ ID NO:4, respectively, would not be expected to retain such biological activities because it is known that this residue in a TGF-β-related polypeptide is required for forming an integral part of the "cysteine knot motif" required for biological activities of the active form of TGF-β family members (McDonald, N. Q. and Hendrickson, W. A. *Cell* **73**:303-304 (1993)).

However, even if deletion of one or more amino acids from the N-terminus of a protein results in modification of loss of one or more biological functions of

the protein, other biological activities may still be retained. Thus, the ability of the shortened proteins to induce and/or bind to antibodies which recognize the complete or mature or active domains of the proteins generally will be retained when less than the majority of the residues of the complete or mature or active domains of the proteins are removed from the N-termini. Whether a particular polypeptide lacking N-terminal residues of a complete protein retains such immunologic activities can readily be determined by routine methods described herein and otherwise known in the art.

Accordingly, the present invention further provides polypeptides having one or more residues deleted from the amino terminus of the amino acid sequence of Nodal shown in SEQ ID NO:2, up to the cysteine residue at position number 183, and polynucleotides encoding such polypeptides. In particular, the present invention provides polypeptides comprising the amino acid sequence of residues n^1-283 of SEQ ID NO:2, where n^1 is an integer in the range of 173-183, and 183 is the position of the first residue from the N-terminus of the complete Nodal polypeptide (shown in SEQ ID NO:2) believed to be required for receptor binding activity of the Nodal protein.

More in particular, the invention provides polynucleotides encoding polypeptides having the amino acid sequence of residues of 173-283, 174-283, 175-283, 176-283, 177-283, 178-283, 179-283, 180-283, 181-283, 182-283, and 183-283 of SEQ ID NO:2. Polynucleotides encoding these polypeptides also are provided.

Further, the present invention also provides polypeptides having one or more residues deleted from the amino terminus of the amino acid sequence of Lefty shown in SEQ ID NO:4, up to the cysteine residue at position number 233, and polynucleotides encoding such polypeptides. In particular, the present invention provides polypeptides comprising the amino acid sequence of residues n^2-348 of SEQ ID NO:4, where n^2 is an integer in the range of 125-233, and 233 is

the position of the first residue from the N-terminus of the complete Nodal polypeptide (shown in SEQ ID NO:4) believed to be required for receptor binding activity of the Lefty protein.

More in particular, the invention provides polynucleotides encoding polypeptides having the amino acid sequence of residues of 125-348, 126-348, 127-348, 128-348, 129-348, 130-348, 131-348, 132-348, 133-348, 134-348, 135-348, 136-348, 137-348, 138-348, 139-348, 140-348, 141-348, 142-348, 143-348, 144-348, 145-348, 146-348, 147-348, 148-348, 149-348, 150-348, 151-348, 152-348, 153-348, 154-348, 155-348, 156-348, 157-348, 158-348, 159-348, 160-348, 161-348, 162-348, 163-348, 164-348, 165-348, 166-348, 167-348, 168-348, 169-348, 170-348, 171-348, 172-348, 173-348, 174-348, 175-348, 176-348, 177-348, 178-348, 179-348, 180-348, 181-348, 182-348, 183-348, 184-348, 185-348, 186-348, 187-348, 188-348, 189-348, 190-348, 191-348, 192-348, 193-348, 194-348, 195-348, 196-348, 197-348, 198-348, 199-348, 200-348, 201-348, 202-348, 203-348, 204-348, 205-348, 206-348, 207-348, 208-348, 209-348, 210-348, 211-348, 212-348, 213-348, 214-348, 215-348, 216-348, 217-348, 218-348, 219-348, 220-348, 221-348, 222-348, 223-348, 224-348, 225-348, 226-348, 227-348, 228-348, 229-348, 230-348, 231-348, 232-348, and 233-348 of SEQ ID NO:4. Polynucleotides encoding these polypeptides also are provided.

Similarly, many examples of biologically functional C-terminal deletion muteins are known. For instance, Interferon gamma shows up to ten times higher activities by deleting 8-10 amino acid residues from the carboxy terminus of the protein (Dobeli, *et al., J. Biotechnology* 7:199-216 (1988)). In the present case, since the proteins of the invention are members of the TGF-β polypeptide family, deletions of C-terminal amino acids up to the cysteine residues at positions 249 and 335 of SEQ ID NO:2 and SEQ ID NO:4, respectively, may retain some biological activity such as receptor binding or modulation of target

cell activities. Polypeptides having further C-terminal deletions including Cys-249 and Cys-335 of SEQ ID NO:2 and SEQ ID NO:4, respectively, would not be expected to retain such biological activities because it is known that this residue in a TGF-β-related polypeptide is required for forming an integral part of the "cysteine knot motif" required for biological activities of the active form of TGF-β family members (McDonald, N. Q. and Hendrickson, W. A. *Cell* **73**:303-304 (1993)).

However, even if deletion of one or more amino acids from the C-terminus of a protein results in modification of loss of one or more biological functions of the protein, other biological activities may still be retained. Thus, the ability of the shortened protein to induce and/or bind to antibodies which recognize the complete, mature or active forms of the protein generally will be retained when less than the majority of the residues of the complete, mature or active forms of the protein are removed from the C-terminus. Whether a particular polypeptide lacking C-terminal residues of a complete protein retains such immunologic activities can readily be determined by routine methods described herein and otherwise known in the art.

Accordingly, the present invention further provides polypeptides having one or more residues from the carboxy terminus of the amino acid sequence of Nodal shown in SEQ ID NO:2, up to the cysteine residue at position 249 of SEQ ID NO:2, and polynucleotides encoding such polypeptides. In particular, the present invention provides polypeptides having the amino acid sequence of residues 1-m^1 of the amino acid sequence in SEQ ID NO:2, where m^1 is any integer in the range of 249 to 283, and residue 249 is the position of the first residue from the C- terminus of the complete Nodal polypeptide (shown in SEQ ID NO:2) believed to be required for receptor binding or modulation of cellular growth and differentiation activities of the Nodal protein.

More in particular, the invention provides polynucleotides encoding polypeptides having the amino acid sequence of residues 1-249, 1-250, 1-251, 1-252, 1-253, 1-254, 1-255, 1-256, 1-257, 1-258, 1-259, 1-260, 1-261, 1-262, 1-263, 1-264, 1-265, 1-266, 1-267, 1-268, 1-269, 1-270, 1-271, 1-272, 1-273, 1-274, 1-275, 1-276, 1-277, 1-278, 1-279, 1-280, 1-281, 1-282, and 1-283 of SEQ ID NO:2. Polynucleotides encoding these polypeptides also are provided.

Further, the present invention also provides polypeptides having one or more residues from the carboxy terminus of the amino acid sequence of Lefty shown in SEQ ID NO:4, up to the cysteine residue at position 335 of SEQ ID NO:4, and polynucleotides encoding such polypeptides. In particular, the present invention provides polypeptides having the amino acid sequence of residues 1-m^2 of the amino acid sequence in SEQ ID NO:4, where m^2 is any integer in the range of 335 to 348, and residue 335 is the position of the first residue from the C-terminus of the complete Lefty polypeptide (shown in SEQ ID NO:4) believed to be required for receptor binding or modulation of cellular growth and differentiation activities of the Lefty protein.

More in particular, the invention provides polynucleotides encoding polypeptides having the amino acid sequence of residues 1-335, 1-336, 1-337, 1-338, 1-339, 1-340, 1-341, 1-342, 1-343, 1-344, 1-345, 1-346, 1-347, and 1-348 of SEQ ID NO:4. Polynucleotides encoding these polypeptides also are provided.

The invention also provides polypeptides having one or more amino acids deleted from both the amino and the carboxyl termini, which may be described generally as having residues n^1-m^1 of SEQ ID NO:2 or n^2-m^2 SEQ ID NO:4, where n^1, m^1, n^2, and m^2 are integers as described above.

Also included is a nucleotide sequence encoding a polypeptide consisting of a portion of the complete Nodal amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 and/or 209135, where this portion

excludes from 1 to about 183 amino acids from the amino terminus of the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 and/or 209135, or from 1 to about 34 amino acids from the carboxy terminus, or any combination of the above amino terminal and carboxy terminal deletions, of the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 and/or 209135.

In addition, a nucleotide sequence encoding a polypeptide consisting of a portion of the complete Lefty amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091 is included, where this portion excludes from 1 to about 250 amino acids from the amino terminus of the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091, or from 1 to about 12 amino acids from the carboxy terminus, or any combination of the above amino terminal and carboxy terminal deletions, of the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091. Polynucleotides encoding all of the above deletion mutant polypeptide forms also are provided.

As mentioned above, even if deletion of one or more amino acids from the N-terminus of a protein results in modification of loss of one or more biological functions of the protein, other biological activities may still be retained. Thus, the ability of the shortened Human Nodal or Human Lefty mutein to induce and/or bind to antibodies which recognize the complete or mature of the protein generally will be retained when less than the majority of the residues of the complete or mature protein are removed from the N-terminus. Whether a particular polypeptide lacking N-terminal residues of a complete protein retains such immunologic activities can readily be determined by routine methods described herein and otherwise known in the art. It is not unlikely that a Human Nodal or Human Lefty mutein with a large number of deleted N-terminal amino acid residues may retain some biological or immungenic activities. In fact,

peptides composed of as few as six Human Nodal or Human Lefty amino acid residues may often evoke an immune response.

Accordingly, the present invention further provides polypeptides having one or more residues deleted from the amino terminus of the Human Nodal amino acid sequence shown in SEQ ID NO:2, up to the glutamic acid residue at position number 278 and polynucleotides encoding such polypeptides. In particular, the present invention provides polypeptides comprising the amino acid sequence of residues n^3-283 of Figures 1A and B (SEQ ID NO:2), where n^3 is an integer in the range of 2 to 278, and 279 is the position of the first residue from the N-terminus of the complete Human Nodal polypeptide believed to be required for at least immunogenic activity of the Human Nodal protein.

More in particular, the invention provides polynucleotides encoding polypeptides comprising, or alternatively consisting of, the amino acid sequence of residues of V-2 to L-283; A-3 to L-283; V-4 to L-283; D-5 to L-283; G-6 to L-283; Q-7 to L-283; N-8 to L-283; W-9 to L-283; T-10 to L-283; F-11 to L-283; A-12 to L-283; F-13 to L-283; D-14 to L-283; F-15 to L-283; S-16 to L-283; F-17 to L-283; L-18 to L-283; S-19 to L-283; Q-20 to L-283; Q-21 to L-283; E-22 to L-283; D-23 to L-283; L-24 to L-283; A-25 to L-283; W-26 to L-283; A-27 to L-283; E-28 to L-283; L-29 to L-283; R-30 to L-283; L-31 to L-283; Q-32 to L-283; L-33 to L-283; S-34 to L-283; S-35 to L-283; P-36 to L-283; V-37 to L-283; D-38 to L-283; L-39 to L-283; P-40 to L-283; T-41 to L-283; E-42 to L-283; G-43 to L-283; S-44 to L-283; L-45 to L-283; A-46 to L-283; I-47 to L-283; E-48 to L-283; I-49 to L-283; F-50 to L-283; H-51 to L-283; Q-52 to L-283; P-53 to L-283; K-54 to L-283; P-55 to L-283; D-56 to L-283; T-57 to L-283; E-58 to L-283; Q-59 to L-283; A-60 to L-283; S-61 to L-283; D-62 to L-283; S-63 to L-283; C-64 to L-283; L-65 to L-283; E-66 to L-283; R-67 to L-283; F-68 to L-283; Q-69 to L-283; M-70 to L-283; D-71 to L-283; L-72 to L-283; F-73 to L-283; T-74 to L-283; V-75 to L-283; T-76 to L-283; L-77 to

L-283; S-78 to L-283; Q-79 to L-283; V-80 to L-283; T-81 to L-283; F-82 to L-283; S-83 to L-283; L-84 to L-283; G-85 to L-283; S-86 to L-283; M-87 to L-283; V-88 to L-283; L-89 to L-283; E-90 to L-283; V-91 to L-283; T-92 to L-283; R-93 to L-283; P-94 to L-283; L-95 to L-283; S-96 to L-283; K-97 to L-283; W-98 to L-283; L-99 to L-283; K-100 to L-283; R-101 to L-283; P-102 to L-283; G-103 to L-283; A-104 to L-283; L-105 to L-283; E-106 to L-283; K-107 to L-283; Q-108 to L-283; M-109 to L-283; S-110 to L-283; R-111 to L-283; V-112 to L-283; A-113 to L-283; G-114 to L-283; E-115 to L-283; C-116 to L-283; W-117 to L-283; P-118 to L-283; R-119 to L-283; P-120 to L-283; P-121 to L-283; T-122 to L-283; P-123 to L-283; P-124 to L-283; A-125 to L-283; T-126 to L-283; N-127 to L-283; V-128 to L-283; L-129 to L-283; L-130 to L-283; M-131 to L-283; L-132 to L-283; Y-133 to L-283; S-134 to L-283; N-135 to L-283; L-136 to L-283; S-137 to L-283; Q-138 to L-283; E-139 to L-283; Q-140 to L-283; R-141 to L-283; Q-142 to L-283; L-143 to L-283; G-144 to L-283; G-145 to L-283; S-146 to L-283; T-147 to L-283; L-148 to L-283; L-149 to L-283; W-150 to L-283; E-151 to L-283; A-152 to L-283; E-153 to L-283; S-154 to L-283; S-155 to L-283; W-156 to L-283; R-157 to L-283; A-158 to L-283; Q-159 to L-283; E-160 to L-283; G-161 to L-283; Q-162 to L-283; L-163 to L-283; S-164 to L-283; W-165 to L-283; E-166 to L-283; W-167 to L-283; G-168 to L-283; K-169 to L-283; R-170 to L-283; H-171 to L-283; R-172 to L-283; R-173 to L-283; H-174 to L-283; H-175 to L-283; L-176 to L-283; P-177 to L-283; D-178 to L-283; R-179 to L-283; S-180 to L-283; Q-181 to L-283; L-182 to L-283; C-183 to L-283; R-184 to L-283; K-185 to L-283; V-186 to L-283; K-187 to L-283; F-188 to L-283; Q-189 to L-283; V-190 to L-283; D-191 to L-283; F-192 to L-283; N-193 to L-283; L-194 to L-283; I-195 to L-283; G-196 to L-283; W-197 to L-283; G-198 to L-283; S-199 to L-283; W-200 to L-283; I-201 to L-283; I-202 to L-283; Y-203 to L-283; P-204 to L-283; K-205 to L-283; Q-206 to L-283; Y-207 to L-283; N-208 to L-283; A-209 to L-283;

Y-210 to L-283; R-211 to L-283; C-212 to L-283; E-213 to L-283; G-214 to L-283; E-215 to L-283; C-216 to L-283; P-217 to L-283; N-218 to L-283; P-219 to L-283; V-220 to L-283; G-221 to L-283; E-222 to L-283; E-223 to L-283; F-224 to L-283; H-225 to L-283; P-226 to L-283; T-227 to L-283; N-228 to L-283; H-229 to L-283; A-230 to L-283; Y-231 to L-283; I-232 to L-283; Q-233 to L-283; S-234 to L-283; L-235 to L-283; L-236 to L-283; K-237 to L-283; R-238 to L-283; Y-239 to L-283; Q-240 to L-283; P-241 to L-283; H-242 to L-283; R-243 to L-283; V-244 to L-283; P-245 to L-283; S-246 to L-283; T-247 to L-283; C-248 to L-283; C-249 to L-283; A-250 to L-283; P-251 to L-283; V-252 to L-283; K-253 to L-283; T-254 to L-283; K-255 to L-283; P-256 to L-283; L-257 to L-283; S-258 to L-283; M-259 to L-283; L-260 to L-283; Y-261 to L-283; V-262 to L-283; D-263 to L-283; N-264 to L-283; G-265 to L-283; R-266 to L-283; V-267 to L-283; L-268 to L-283; L-269 to L-283; D-270 to L-283; H-271 to L-283; H-272 to L-283; K-273 to L-283; D-274 to L-283; M-275 to L-283; I-276 to L-283; V-277 to L-283; and E-278 to L-283 of the Human Nodal sequence shown in Figures 1A and B (which is identical to the Human Nodal sequence in SEQ ID NO:2). Polynucleotides encoding these polypeptides are also encompassed by the invention.

Also as mentioned above, even if deletion of one or more amino acids from the C-terminus of a protein results in modification of loss of one or more biological functions of the protein, other biological activities may still be retained. Thus, the ability of the shortened Human Nodal mutein to induce and/or bind to antibodies which recognize the complete or mature of the protein generally will be retained when less than the majority of the residues of the complete or mature protein are removed from the C-terminus. Whether a particular polypeptide lacking C-terminal residues of a complete protein retains such immunologic activities can readily be determined by routine methods described herein and otherwise known in the art. It is not unlikely that a Human Nodal mutein with a

large number of deleted C-terminal amino acid residues may retain some biological or immungenic activities. In fact, peptides composed of as few as six Human Nodal amino acid residues may often evoke an immune response.

Accordingly, the present invention further provides polypeptides having one or more residues deleted from the carboxy terminus of the amino acid sequence of the Human Nodal shown in SEQ ID NO:2, up to the glycine residue at position number 6, and polynucleotides encoding such polypeptides. In particular, the present invention provides polypeptides comprising the amino acid sequence of residues 1-m^3 of SEQ ID NO:2, where m^3 is an integer in the range of 6 to 283, and 6 is the position of the first residue from the C-terminus of the complete Human Nodal polypeptide believed to be required for at least immunogenic activity of the Human Nodal protein.

More in particular, the invention provides polynucleotides encoding polypeptides comprising, or alternatively consisting of, the amino acid sequence of residues D-1 to C-282; D-1 to G-281; D-1 to C-280; D-1 to E-279; D-1 to E-278; D-1 to V-277; D-1 to I-276; D-1 to M-275; D-1 to D-274; D-1 to K-273; D-1 to H-272; D-1 to H-271; D-1 to D-270; D-1 to L-269; D-1 to L-268; D-1 to V-267; D-1 to R-266; D-1 to G-265; D-1 to N-264; D-1 to D-263; D-1 to V-262; D-1 to Y-261; D-1 to L-260; D-1 to M-259; D-1 to S-258; D-1 to L-257; D-1 to P-256; D-1 to K-255; D-1 to T-254; D-1 to K-253; D-1 to V-252; D-1 to P-251; D-1 to A-250; D-1 to C-249; D-1 to C-248; D-1 to T-247; D-1 to S-246; D-1 to P-245; D-1 to V-244; D-1 to R-243; D-1 to H-242; D-1 to P-241; D-1 to Q-240; D-1 to Y-239; D-1 to R-238; D-1 to K-237; D-1 to L-236; D-1 to L-235; D-1 to S-234; D-1 to Q-233; D-1 to I-232; D-1 to Y-231; D-1 to A-230; D-1 to H-229; D-1 to N-228; D-1 to T-227; D-1 to P-226; D-1 to H-225; D-1 to F-224; D-1 to E-223; D-1 to E-222; D-1 to G-221; D-1 to V-220; D-1 to P-219; D-1 to N-218; D-1 to P-217; D-1 to C-216; D-1 to E-215; D-1 to G-214; D-1 to E-213; D-1 to C-212; D-1 to R-211; D-1 to Y-210; D-1 to A-209; D-1 to N-208; D-1 to Y-207;

D-1 to Q-206; D-1 to K-205; D-1 to P-204; D-1 to Y-203; D-1 to I-202; D-1 to I-201; D-1 to W-200; D-1 to S-199; D-1 to G-198; D-1 to W-197; D-1 to G-196; D-1 to I-195; D-1 to L-194; D-1 to N-193; D-1 to F-192; D-1 to D-191; D-1 to V-190; D-1 to Q-189; D-1 to F-188; D-1 to K-187; D-1 to V-186; D-1 to K-185; D-1 to R-184; D-1 to C-183; D-1 to L-182; D-1 to Q-181; D-1 to S-180; D-1 to R-179; D-1 to D-178; D-1 to P-177; D-1 to L-176; D-1 to H-175; D-1 to H-174; D-1 to R-173; D-1 to R-172; D-1 to H-171; D-1 to R-170; D-1 to K-169; D-1 to G-168; D-1 to W-167; D-1 to E-166; D-1 to W-165; D-1 to S-164; D-1 to L-163; D-1 to Q-162; D-1 to G-161; D-1 to E-160; D-1 to Q-159; D-1 to A-158; D-1 to R-157; D-1 to W-156; D-1 to S-155; D-1 to S-154; D-1 to E-153; D-1 to A-152; D-1 to E-151; D-1 to W-150; D-1 to L-149; D-1 to L-148; D-1 to T-147; D-1 to S-146; D-1 to G-145; D-1 to G-144; D-1 to L-143; D-1 to Q-142; D-1 to R-141; D-1 to Q-140; D-1 to E-139; D-1 to Q-138; D-1 to S-137; D-1 to L-136; D-1 to N-135; D-1 to S-134; D-1 to Y-133; D-1 to L-132; D-1 to M-131; D-1 to L-130; D-1 to L-129; D-1 to V-128; D-1 to N-127; D-1 to T-126; D-1 to A-125; D-1 to P-124; D-1 to P-123; D-1 to T-122; D-1 to P-121; D-1 to P-120; D-1 to R-119; D-1 to P-118; D-1 to W-117; D-1 to C-116; D-1 to E-115; D-1 to G-114; D-1 to A-113; D-1 to V-112; D-1 to R-111; D-1 to S-110; D-1 to M-109; D-1 to Q-108; D-1 to K-107; D-1 to E-106; D-1 to L-105; D-1 to A-104; D-1 to G-103; D-1 to P-102; D-1 to R-101; D-1 to K-100; D-1 to L-99; D-1 to W-98; D-1 to K-97; D-1 to S-96; D-1 to L-95; D-1 to P-94; D-1 to R-93; D-1 to T-92; D-1 to V-91; D-1 to E-90; D-1 to L-89; D-1 to V-88; D-1 to M-87; D-1 to S-86; D-1 to G-85; D-1 to L-84; D-1 to S-83; D-1 to F-82; D-1 to T-81; D-1 to V-80; D-1 to Q-79; D-1 to S-78; D-1 to L-77; D-1 to T-76; D-1 to V-75; D-1 to T-74; D-1 to F-73; D-1 to L-72; D-1 to D-71; D-1 to M-70; D-1 to Q-69; D-1 to F-68; D-1 to R-67; D-1 to E-66; D-1 to L-65; D-1 to C-64; D-1 to S-63; D-1 to D-62; D-1 to S-61; D-1 to A-60; D-1 to Q-59; D-1 to E-58; D-1 to T-57; D-1 to D-56; D-1 to P-55; D-1 to K-54; D-1 to P-53; D-1 to Q-52; D-1 to H-51; D-1 to F-50; D-1 to I-49;

D-1 to E-48; D-1 to I-47; D-1 to A-46; D-1 to L-45; D-1 to S-44; D-1 to G-43; D-1 to E-42; D-1 to T-41; D-1 to P-40; D-1 to L-39; D-1 to D-38; D-1 to V-37; D-1 to P-36; D-1 to S-35; D-1 to S-34; D-1 to L-33; D-1 to Q-32; D-1 to L-31; D-1 to R-30; D-1 to L-29; D-1 to E-28; D-1 to A-27; D-1 to W-26; D-1 to A-25; D-1 to L-24; D-1 to D-23; D-1 to E-22; D-1 to Q-21; D-1 to Q-20; D-1 to S-19; D-1 to L-18; D-1 to F-17; D-1 to S-16; D-1 to F-15; D-1 to D-14; D-1 to F-13; D-1 to A-12; D-1 to F-11; D-1 to T-10; D-1 to W-9; D-1 to N-8; D-1 to Q-7; D-1 to G-6 of the sequence of the Human Nodal sequence shown in Figures 1A and B (which is identical to the Human Nodal sequence shown in SEQ ID NO:2). Polynucleotides encoding these polypeptides also are provided.

The invention also provides polypeptides having one or more amino acids deleted from both the amino and the carboxyl termini of a Human Nodal polypeptide, which may be described generally as having residues n^3-m^3 of Figures 1A and B (SEQ ID NO:2), where n^3 and m^3 are integers as described above.

Again as mentioned above, even if deletion of one or more amino acids from the N-terminus of a protein results in modification of loss of one or more biological functions of the protein, other biological activities may still be retained. Thus, the ability of the shortened Human Lefty mutein to induce and/or bind to antibodies which recognize the complete or mature of the protein generally will be retained when less than the majority of the residues of the complete or mature protein are removed from the N-terminus. Whether a particular polypeptide lacking N-terminal residues of a complete protein retains such immunologic activities can readily be determined by routine methods described herein and otherwise known in the art. It is not unlikely that a Human Lefty mutein with a large number of deleted N-terminal amino acid residues may retain some biological or immungenic activities. In fact, peptides composed of as few as six Human Lefty amino acid residues may often evoke an immune response.

Accordingly, the present invention further provides polypeptides having one or more residues deleted from the amino terminus of the Human Lefty amino acid sequence shown in SEQ ID NO:4, up to the proline residue at position number 361 and polynucleotides encoding such polypeptides. In particular, the present invention provides polypeptides comprising the amino acid sequence of residues n^4-180 of Figures 2A and B (SEQ ID NO:4), where n^4 is an integer in the range of 2 to 361, and 362 is the position of the first residue from the N-terminus of the complete Human Lefty polypeptide believed to be required for at least immunogenic activity of the Human Lefty protein.

More in particular, the invention provides polynucleotides encoding polypeptides comprising, or alternatively consisting of, the amino acid sequence of residues of Q-2 to P-366; P-3 to P-366; L-4 to P-366; W-5 to P-366; L-6 to P-366; C-7 to P-366; W-8 to P-366; A-9 to P-366; L-10 to P-366; W-11 to P-366; V-12 to P-366; L-13 to P-366; P-14 to P-366; L-15 to P-366; A-16 to P-366; S-17 to P-366; P-18 to P-366; G-19 to P-366; A-20 to P-366; A-21 to P-366; L-22 to P-366; T-23 to P-366; G-24 to P-366; E-25 to P-366; Q-26 to P-366; L-27 to P-366; L-28 to P-366; G-29 to P-366; S-30 to P-366; L-31 to P-366; L-32 to P-366; R-33 to P-366; Q-34 to P-366; L-35 to P-366; Q-36 to P-366; L-37 to P-366; K-38 to P-366; E-39 to P-366; V-40 to P-366; P-41 to P-366; T-42 to P-366; L-43 to P-366; D-44 to P-366; R-45 to P-366; A-46 to P-366; D-47 to P-366; M-48 to P-366; E-49 to P-366; E-50 to P-366; L-51 to P-366; V-52 to P-366; I-53 to P-366; P-54 to P-366; T-55 to P-366; H-56 to P-366; V-57 to P-366; R-58 to P-366; A-59 to P-366; Q-60 to P-366; Y-61 to P-366; V-62 to P-366; A-63 to P-366; L-64 to P-366; L-65 to P-366; Q-66 to P-366; R-67 to P-366; S-68 to P-366; H-69 to P-366; G-70 to P-366; D-71 to P-366; R-72 to P-366; S-73 to P-366; R-74 to P-366; G-75 to P-366; K-76 to P-366; R-77 to P-366; F-78 to P-366; S-79 to P-366; Q-80 to P-366; S-81 to P-366; F-82 to P-366; R-83 to P-366; E-84 to P-366; V-85 to P-366; A-86 to

P-366; G-87 to P-366; R-88 to P-366; F-89 to P-366; L-90 to P-366; A-91 to P-366; L-92 to P-366; E-93 to P-366; A-94 to P-366; S-95 to P-366; T-96 to P-366; H-97 to P-366; L-98 to P-366; L-99 to P-366; V-100 to P-366; F-101 to P-366; G-102 to P-366; M-103 to P-366; E-104 to P-366; Q-105 to P-366; R-106 to P-366; L-107 to P-366; P-108 to P-366; P-109 to P-366; N-110 to P-366; S-111 to P-366; E-112 to P-366; L-113 to P-366; V-114 to P-366; Q-115 to P-366; A-116 to P-366; V-117 to P-366; L-118 to P-366; R-119 to P-366; L-120 to P-366; F-121 to P-366; Q-122 to P-366; E-123 to P-366; P-124 to P-366; V-125 to P-366; P-126 to P-366; K-127 to P-366; A-128 to P-366; A-129 to P-366; L-130 to P-366; H-131 to P-366; R-132 to P-366; H-133 to P-366; G-134 to P-366; R-135 to P-366; L-136 to P-366; S-137 to P-366; P-138 to P-366; R-139 to P-366; S-140 to P-366; A-141 to P-366; R-142 to P-366; A-143 to P-366; R-144 to P-366; V-145 to P-366; T-146 to P-366; V-147 to P-366; E-148 to P-366; W-149 to P-366; L-150 to P-366; R-151 to P-366; V-152 to P-366; R-153 to P-366; D-154 to P-366; D-155 to P-366; G-156 to P-366; S-157 to P-366; N-158 to P-366; R-159 to P-366; T-160 to P-366; S-161 to P-366; L-162 to P-366; I-163 to P-366; D-164 to P-366; S-165 to P-366; R-166 to P-366; L-167 to P-366; V-168 to P-366; S-169 to P-366; V-170 to P-366; H-171 to P-366; E-172 to P-366; S-173 to P-366; G-174 to P-366; W-175 to P-366; K-176 to P-366; A-177 to P-366; F-178 to P-366; D-179 to P-366; V-180 to P-366; T-181 to P-366; E-182 to P-366; A-183 to P-366; V-184 to P-366; N-185 to P-366; F-186 to P-366; W-187 to P-366; Q-188 to P-366; Q-189 to P-366; L-190 to P-366; S-191 to P-366; R-192 to P-366; P-193 to P-366; R-194 to P-366; Q-195 to P-366; P-196 to P-366; L-197 to P-366; L-198 to P-366; L-199 to P-366; Q-200 to P-366; V-201 to P-366; S-202 to P-366; V-203 to P-366; Q-204 to P-366; R-205 to P-366; E-206 to P-366; H-207 to P-366; L-208 to P-366; G-209 to P-366; P-210 to P-366; L-211 to P-366; A-212 to P-366; S-213 to P-366; G-214 to P-366; A-215 to P-366; H-216 to P-366; K-217 to P-366;

L-218 to P-366; V-219 to P-366; R-220 to P-366; F-221 to P-366; A-222 to P-366; S-223 to P-366; Q-224 to P-366; G-225 to P-366; A-226 to P-366; P-227 to P-366; A-228 to P-366; G-229 to P-366; L-230 to P-366; G-231 to P-366; E-232 to P-366; P-233 to P-366; Q-234 to P-366; L-235 to P-366; E-236 to P-366; L-237 to P-366; H-238 to P-366; T-239 to P-366; L-240 to P-366; D-241 to P-366; L-242 to P-366; G-243 to P-366; D-244 to P-366; Y-245 to P-366; G-246 to P-366; A-247 to P-366; Q-248 to P-366; G-249 to P-366; D-250 to P-366; C-251 to P-366; D-252 to P-366; P-253 to P-366; E-254 to P-366; A-255 to P-366; P-256 to P-366; M-257 to P-366; T-258 to P-366; E-259 to P-366; G-260 to P-366; T-261 to P-366; R-262 to P-366; C-263 to P-366; C-264 to P-366; R-265 to P-366; Q-266 to P-366; E-267 to P-366; M-268 to P-366; Y-269 to P-366; I-270 to P-366; D-271 to P-366; L-272 to P-366; Q-273 to P-366; G-274 to P-366; M-275 to P-366; K-276 to P-366; W-277 to P-366; A-278 to P-366; E-279 to P-366; N-280 to P-366; W-281 to P-366; V-282 to P-366; L-283 to P-366; E-284 to P-366; P-285 to P-366; P-286 to P-366; G-287 to P-366; F-288 to P-366; L-289 to P-366; A-290 to P-366; Y-291 to P-366; E-292 to P-366; C-293 to P-366; V-294 to P-366; G-295 to P-366; T-296 to P-366; C-297 to P-366; R-298 to P-366; Q-299 to P-366; P-300 to P-366; P-301 to P-366; E-302 to P-366; A-303 to P-366; L-304 to P-366; A-305 to P-366; F-306 to P-366; K-307 to P-366; W-308 to P-366; P-309 to P-366; F-310 to P-366; L-311 to P-366; G-312 to P-366; P-313 to P-366; R-314 to P-366; Q-315 to P-366; C-316 to P-366; I-317 to P-366; A-318 to P-366; S-319 to P-366; E-320 to P-366; T-321 to P-366; D-322 to P-366; S-323 to P-366; L-324 to P-366; P-325 to P-366; M-326 to P-366; I-327 to P-366; V-328 to P-366; S-329 to P-366; I-330 to P-366; K-331 to P-366; E-332 to P-366; G-333 to P-366; G-334 to P-366; R-335 to P-366; T-336 to P-366; R-337 to P-366; P-338 to P-366; Q-339 to P-366; V-340 to P-366; V-341 to P-366; S-342 to P-366; L-343 to P-366; P-344 to P-366; N-345 to P-366; M-346 to P-366; R-347 to P-366; V-348 to

P-366; Q-349 to P-366; K-350 to P-366; C-351 to P-366; S-352 to P-366; C-353 to P-366; A-354 to P-366; S-355 to P-366; D-356 to P-366; G-357 to P-366; A-358 to P-366; L-359 to P-366; V-360 to P-366; and P-361 to P-366 of the Human Lefty sequence shown in Figures 2A and B (which is identical to the sequence shown as SEQ ID NO:4, with the exception that the amino acid residues in Figures 2A and B are numbered consecutively from 1 through 366 from the N-terminus to the C-terminus, while the amino acid residues in SEQ ID NO:4 are numbered consecutively from -18 through 348 to reflect the position of the predicted signal peptide). Polynucleotides encoding these polypeptides are also encompassed by the invention.

Also as mentioned above, even if deletion of one or more amino acids from the C-terminus of a protein results in modification of loss of one or more biological functions of the protein, other biological activities may still be retained. Thus, the ability of the shortened Human Lefty mutein to induce and/or bind to antibodies which recognize the complete or mature of the protein generally will be retained when less than the majority of the residues of the complete or mature protein are removed from the C-terminus. Whether a particular polypeptide lacking C-terminal residues of a complete protein retains such immunologic activities can readily be determined by routine methods described herein and otherwise known in the art. It is not unlikely that a Human Lefty mutein with a large number of deleted C-terminal amino acid residues may retain some biological or immungenic activities. In fact, peptides composed of as few as six Human Lefty amino acid residues may often evoke an immune response.

Accordingly, the present invention further provides polypeptides having one or more residues deleted from the carboxy terminus of the amino acid sequence of the Human Lefty shown in SEQ ID NO:4, up to the leucine residue at position number 6, and polynucleotides encoding such polypeptides. In particular, the present invention provides polypeptides comprising the amino

acid sequence of residues 1-m^4 of SEQ ID NO:4, where m^4 is an integer in the range of 6 to 366, and 6 is the position of the first residue from the C-terminus of the complete Human Lefty polypeptide believed to be required for at least immunogenic activity of the Human Lefty protein.

More in particular, the invention provides polynucleotides encoding polypeptides comprising, or alternatively consisting of, the amino acid sequence of residues M-1 to Q-365; M-1 to L-364; M-1 to R-363; M-1 to R-362; M-1 to P-361; M-1 to V-360; M-1 to L-359; M-1 to A-358; M-1 to G-357; M-1 to D-356; M-1 to S-355; M-1 to A-354; M-1 to C-353; M-1 to S-352; M-1 to C-351; M-1 to K-350; M-1 to Q-349; M-1 to V-348; M-1 to R-347; M-1 to M-346; M-1 to N-345; M-1 to P-344; M-1 to L-343; M-1 to S-342; M-1 to V-341; M-1 to V-340; M-1 to Q-339; M-1 to P-338; M-1 to R-337; M-1 to T-336; M-1 to R-335; M-1 to G-334; M-1 to G-333; M-1 to E-332; M-1 to K-331; M-1 to I-330; M-1 to S-329; M-1 to V-328; M-1 to I-327; M-1 to M-326; M-1 to P-325; M-1 to L-324; M-1 to S-323; M-1 to D-322; M-1 to T-321; M-1 to E-320; M-1 to S-319; M-1 to A-318; M-1 to I-317; M-1 to C-316; M-1 to Q-315; M-1 to R-314; M-1 to P-313; M-1 to G-312; M-1 to L-311; M-1 to F-310; M-1 to P-309; M-1 to W-308; M-1 to K-307; M-1 to F-306; M-1 to A-305; M-1 to L-304; M-1 to A-303; M-1 to E-302; M-1 to P-301; M-1 to P-300; M-1 to Q-299; M-1 to R-298; M-1 to C-297; M-1 to T-296; M-1 to G-295; M-1 to V-294; M-1 to C-293; M-1 to E-292; M-1 to Y-291; M-1 to A-290; M-1 to L-289; M-1 to F-288; M-1 to G-287; M-1 to P-286; M-1 to P-285; M-1 to E-284; M-1 to L-283; M-1 to V-282; M-1 to W-281; M-1 to N-280; M-1 to E-279; M-1 to A-278; M-1 to W-277; M-1 to K-276; M-1 to M-275; M-1 to G-274; M-1 to Q-273; M-1 to L-272; M-1 to D-271; M-1 to I-270; M-1 to Y-269; M-1 to M-268; M-1 to E-267; M-1 to Q-266; M-1 to R-265; M-1 to C-264; M-1 to C-263; M-1 to R-262; M-1 to T-261; M-1 to G-260; M-1 to E-259; M-1 to T-258; M-1 to M-257; M-1 to

P-256; M-1 to A-255; M-1 to E-254; M-1 to P-253; M-1 to D-252; M-1 to C-251; M-1 to D-250; M-1 to G-249; M-1 to Q-248; M-1 to A-247; M-1 to G-246; M-1 to Y-245; M-1 to D-244; M-1 to G-243; M-1 to L-242; M-1 to D-241; M-1 to L-240; M-1 to T-239; M-1 to H-238; M-1 to L-237; M-1 to E-236; M-1 to L-235; M-1 to Q-234; M-1 to P-233; M-1 to E-232; M-1 to G-231; M-1 to L-230; M-1 to G-229; M-1 to A-228; M-1 to P-227; M-1 to A-226; M-1 to G-225; M-1 to Q-224; M-1 to S-223; M-1 to A-222; M-1 to F-221; M-1 to R-220; M-1 to V-219; M-1 to L-218; M-1 to K-217; M-1 to H-216; M-1 to A-215; M-1 to G-214; M-1 to S-213; M-1 to A-212; M-1 to L-211; M-1 to P-210; M-1 to G-209; M-1 to L-208; M-1 to H-207; M-1 to E-206; M-1 to R-205; M-1 to Q-204; M-1 to V-203; M-1 to S-202; M-1 to V-201; M-1 to Q-200; M-1 to L-199; M-1 to L-198; M-1 to L-197; M-1 to P-196; M-1 to Q-195; M-1 to R-194; M-1 to P-193; M-1 to R-192; M-1 to S-191; M-1 to L-190; M-1 to Q-189; M-1 to Q-188; M-1 to W-187; M-1 to F-186; M-1 to N-185; M-1 to V-184; M-1 to A-183; M-1 to E-182; M-1 to T-181; M-1 to V-180; M-1 to D-179; M-1 to F-178; M-1 to A-177; M-1 to K-176; M-1 to W-175; M-1 to G-174; M-1 to S-173; M-1 to E-172; M-1 to H-171; M-1 to V-170; M-1 to S-169; M-1 to V-168; M-1 to L-167; M-1 to R-166; M-1 to S-165; M-1 to D-164; M-1 to I-163; M-1 to L-162; M-1 to S-161; M-1 to T-160; M-1 to R-159; M-1 to N-158; M-1 to S-157; M-1 to G-156; M-1 to D-155; M-1 to D-154; M-1 to R-153; M-1 to V-152; M-1 to R-151; M-1 to L-150; M-1 to W-149; M-1 to E-148; M-1 to V-147; M-1 to T-146; M-1 to V-145; M-1 to R-144; M-1 to A-143; M-1 to R-142; M-1 to A-141; M-1 to S-140; M-1 to R-139; M-1 to P-138; M-1 to S-137; M-1 to L-136; M-1 to R-135; M-1 to G-134; M-1 to H-133; M-1 to R-132; M-1 to H-131; M-1 to L-130; M-1 to A-129; M-1 to A-128; M-1 to K-127; M-1 to P-126; M-1 to V-125; M-1 to P-124; M-1 to E-123; M-1 to Q-122; M-1 to F-121; M-1 to L-120; M-1 to R-119; M-1 to L-118; M-1 to V-117; M-1 to

A-116; M-1 to Q-115; M-1 to V-114; M-1 to L-113; M-1 to E-112; M-1 to S-111; M-1 to N-110; M-1 to P-109; M-1 to P-108; M-1 to L-107; M-1 to R-106; M-1 to Q-105; M-1 to E-104; M-1 to M-103; M-1 to G-102; M-1 to F-101; M-1 to V-100; M-1 to L-99; M-1 to L-98; M-1 to H-97; M-1 to T-96; M-1 to S-95; M-1 to A-94; M-1 to E-93; M-1 to L-92; M-1 to A-91; M-1 to L-90; M-1 to F-89; M-1 to R-88; M-1 to G-87; M-1 to A-86; M-1 to V-85; M-1 to E-84; M-1 to R-83; M-1 to F-82; M-1 to S-81; M-1 to Q-80; M-1 to S-79; M-1 to F-78; M-1 to R-77; M-1 to K-76; M-1 to G-75; M-1 to R-74; M-1 to S-73; M-1 to R-72; M-1 to D-71; M-1 to G-70; M-1 to H-69; M-1 to S-68; M-1 to R-67; M-1 to Q-66; M-1 to L-65; M-1 to L-64; M-1 to A-63; M-1 to V-62; M-1 to Y-61; M-1 to Q-60; M-1 to A-59; M-1 to R-58; M-1 to V-57; M-1 to H-56; M-1 to T-55; M-1 to P-54; M-1 to I-53; M-1 to V-52; M-1 to L-51; M-1 to E-50; M-1 to E-49; M-1 to M-48; M-1 to D-47; M-1 to A-46; M-1 to R-45; M-1 to D-44; M-1 to L-43; M-1 to T-42; M-1 to P-41; M-1 to V-40; M-1 to E-39; M-1 to K-38; M-1 to L-37; M-1 to Q-36; M-1 to L-35; M-1 to Q-34; M-1 to R-33; M-1 to L-32; M-1 to L-31; M-1 to S-30; M-1 to G-29; M-1 to L-28; M-1 to L-27; M-1 to Q-26; M-1 to E-25; M-1 to G-24; M-1 to T-23; M-1 to L-22; M-1 to A-21; M-1 to A-20; M-1 to G-19; M-1 to P-18; M-1 to S-17; M-1 to A-16; M-1 to L-15; M-1 to P-14; M-1 to L-13; M-1 to V-12; M-1 to W-11; M-1 to L-10; M-1 to A-9; M-1 to W-8; M-1 to C-7; and M-1 to L-6 of the sequence of the Human Lefty sequence shown in Figures 2A and B (which is identical to the sequence shown as SEQ ID NO:4, with the exception that the amino acid residues in Figures 2A and B are numbered consecutively from 1 through 366 from the N-terminus to the C-terminus, while the amino acid residues in SEQ ID NO:4 are numbered consecutively from -18 through 348 to reflect the position of the predicted signal peptide). Polynucleotides encoding these polypeptides also are provided.

The invention also provides polypeptides having one or more amino acids deleted from both the amino and the carboxyl termini of a Human Lefty polypeptide, which may be described generally as having residues n^4-m^4 of Figures 2A and B (SEQ ID NO:4), where n^4 and m^4 are integers as described above.

In addition to terminal deletion forms of the proteins discussed above, it also will be recognized by one of ordinary skill in the art that some amino acid sequences of the Nodal and Lefty polypeptides can be varied without significant effect of the structure or function of the proteins. If such differences in sequence are contemplated, it should be remembered that there will be critical areas on the protein which determine activity.

Thus, the invention further includes variations of the Nodal and Lefty polypeptides which show substantial Nodal or Lefty polypeptide activity or which include regions of Nodal or Lefty proteins such as the protein portions discussed below. Such mutants include deletions, insertions, inversions, repeats, and type substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided wherein the authors indicate that there are two main approaches for studying the tolerance of an amino acid sequence to change (Bowie, J. U., *et al.*, *Science* **247**:1306-1310 (1990)),. The first method relies on the process of evolution, in which mutations are either accepted or rejected by natural selection. The second approach uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene and selections or screens to identify sequences that maintain functionality.

As the authors state, these studies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at a certain position of the

protein. For example, most buried amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Other such phenotypically silent substitutions are described by Bowie and coworkers (*supra*) and the references cited therein. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

Thus, the fragment, derivative or analog of the polypeptides of SEQ ID NO:2 or SEQ ID NO:4, or those encoded by the deposited cDNAs, may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the active form of the polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the above form of the polypeptide, such as an IgG Fc fusion region peptide or leader or secretory sequence or a sequence which is employed for purification of the above form of the polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

Thus, the Nodal and Lefty proteins of the present invention may include one or more amino acid substitutions, deletions or additions, either from natural mutations or human manipulation. As indicated, changes are preferably of a

minor nature, such as conservative amino acid substitutions that do not significantly affect the folding or activity of the protein (see Table 1).

TABLE 1. Conservative Amino Acid Substitutions.

Aromatic	Phenylalanine Tryptophan Tyrosine
Hydrophobic	Leucine Isoleucine Valine
Polar	Glutamine Asparagine
Basic	Arginine Lysine Histidine
Acidic	Aspartic Acid Glutamic Acid
Small	Alanine Serine Threonine Methionine Glycine

Embodiments of the invention are directed to polypeptides which comprise the amino acid sequence of a Nodal or Lefty polypeptide described herein, but having an amino acid sequence which contains at least one conservative amino acid substitution, but not more than 50 conservative amino acid substitutions, even more preferably, not more than 40 conservative amino acid substitutions, still more preferably, not more than 30 conservative amino acid substitutions, and still even more preferably, not more than 20 conservative

amino acid substitutions, when compared with the Nodal or Lefty polynucleotide sequence described herein. Of course, in order of ever-increasing preference, it is highly preferable for a peptide or polypeptide to have an amino acid sequence which comprises the amino acid sequence of a Nodal or Lefty polypeptide, which contains at least one, but not more than 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 conservative amino acid substitutions.

In further specific embodiments, the number of substitutions, additions or deletions in the amino acid sequence of Figures 1A and B (SEQ ID NO:2), Figures 2A and B (SEQ ID NO:4), a polypeptide sequence encoded by the deposited clones, and/or any of the polypeptide fragments described herein (e.g., the mature forms or the active TGF-β consensus cleavage domains) is 75, 70, 60, 50, 40, 35, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1 or 150-50, 100-50, 50-20, 30-20, 20-15, 20-10, 15-10, 10-1, 5-10, 1-5, 1-3 or 1-2.

To improve or alter the characteristics of Nodal or Lefty polypeptides, protein engineering may be employed. Recombinant DNA technology known to those skilled in the art can be used to create novel mutant polypeptides or muteins including single or multiple amino acid substitutions, deletions, additions or fusion proteins. Such modified polypeptides can show, e.g., enhanced activity or increased stability. In addition, they may be purified in higher yields and show better solubility than the corresponding natural polypeptide, at least under certain purification and storage conditions.

Thus, the invention also encompasses Nodal and Lefty derivatives and analogs that have one or more amino acid residues deleted, added, or substituted to generate Nodal and Lefty polypeptides that are better suited for expression, scale up, etc., in the host cells chosen. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges; N-linked glycosylation sites can be altered or eliminated to achieve, for example, expression of a homogeneous product that is more easily

recovered and purified from yeast hosts which are known to hyperglycosylate N-linked sites. To this end, a variety of amino acid substitutions at one or both of the first or third amino acid positions on any one or more of the glycosylation recognition sequences in the Nodal and Lefty polypeptides of the invention, and/or an amino acid deletion at the second position of any one or more such recognition sequences will prevent glycosylation of the Nodal or Lefty polypeptide at the modified tripeptide sequence (see, e.g., Miyajima, A., *et al.*, *EMBO J.* 5(6):1193-1197 (1986)).

Amino acids in the Nodal and Lefty polypeptides of the present invention that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, *Science* **244:**1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as receptor binding or *in vitro* proliferative activity.

Of special interest are substitutions of charged amino acids with other charged or neutral amino acids which may produce proteins with highly desirable improved characteristics, such as less aggregation. Aggregation may not only reduce activity but also be problematic when preparing pharmaceutical formulations, because aggregates can be immunogenic (Pinckard, *et al., Clin. Exp. Immunol.* **2:**331-340 (1967); Robbins, *et al., Diabetes* **36:**838-845 (1987); Cleland, *et al., Crit. Rev. Therapeutic Drug Carrier Systems* **10:**307-377 (1993)).

Replacement of amino acids can also change the selectivity of the binding of a ligand to cell surface receptors (for example, Ostade, *et al., Nature* **361:**266-268 (1993)) describes certain mutations resulting in selective binding of TNF-α to only one of the two known types of TNF receptors. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling

(Smith, *et al., J. Mol. Biol.* **224**:899-904 (1992); de Vos, *et al. Science* **255**:306-312 (1992)).

Since Nodal and Lefty are members of the TGF-β-related protein family, to modulate rather than completely eliminate biological activities of Nodal and Lefty preferably mutations are made in sequences encoding amino acids in the Nodal and Lefty conserved domain, i.e., in positions 173 to 283 or SEQ ID NO:2 or positions 125 to 348 of SEQ ID NO:4, more preferably in residues within this region which are not conserved in all members of the TGF-β-related protein family. In particular, mutations to the Nodal and Lefty polypeptides are mad in positions other than the conserved cysteine residues comprising the "cysteine knot" motif characteristic of TGF-β-related protein family members. Also forming part of the present invention are isolated polynucleotides comprising nucleic acid sequences which encode the above Nodal and Lefty mutants.

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. Recombinantly produced versions of the Nodal and Lefty polypeptides can be substantially purified by the one-step method described by Smith and Johnson (*Gene* **67**:31-40 (1988)). Polypeptides of the invention also can be purified from natural or recombinant sources using anti-Nodal or anti-Lefty antibodies of the invention in methods which are well known in the art of protein purification.

The invention further provides isolated Nodal and Lefty polypeptides comprising an amino acid sequence selected from the group consisting of: (a) the amino acid sequence of the full-length Nodal polypeptide having the complete amino acid sequence shown in SEQ ID NO:2 (i.e., positions 1 to 283 of SEQ ID NO:2); (b) the amino acid sequence of the predicted active Nodal polypeptide having the amino acid sequence at positions 173 to 283 of SEQ ID NO:2; (c) the amino acid sequence of the Nodal polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092

and/or 209135; (d) the amino acid sequence of the active domain of the Nodal polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 and/or 209135; (e) the amino acid sequence of the Lefty polypeptide having the complete amino acid sequence in SEQ ID NO:4 (i.e., positions -18 to 348 of SEQ ID NO:4); (f) the amino acid sequence of the Lefty polypeptide having the complete amino acid sequence in SEQ ID NO:4 excepting the N-terminal methionine (i.e., positions -17 to 348 of SEQ ID NO:4); (g) the amino acid sequence of the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 60 to 348 of SEQ ID NO:4; (h) the amino acid sequence of the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 118 to 348 of SEQ ID NO:4; (i) the amino acid sequence of the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 125 to 348 of SEQ ID NO:4; (j) the amino acid sequence of the Lefty polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091; (k) the amino acid sequence of the Lefty polypeptide having the complete amino acid sequence excepting the N-terminal methionine encoded by the cDNA clone contained in ATCC Deposit No. 209091, and; (l) the amino acid sequence of the active domain of the Lefty polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091.

Further polypeptides of the present invention include polypeptides which have at least 90% similarity, more preferably at least 95% similarity, and still more preferably at least 96%, 97%, 98% or 99% similarity to those described above. The polypeptides of the invention also comprise those which are at least 80% identical, more preferably at least 90% or 95% identical, still more preferably at least 96%, 97%, 98% or 99% identical to the polypeptide encoded by the deposited cDNAs or to the polypeptides of SEQ ID NO:2 or SEQ ID

NO:4, and also include portions of such polypeptides with at least 30 amino acids and more preferably at least 50 amino acids.

By "% similarity" for two polypeptides is intended a similarity score produced by comparing the amino acid sequences of the two polypeptides using the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711) and the default settings for determining similarity. Bestfit uses the local homology algorithm of Smith and Waterman (Advances in Applied Mathematics 2:482-489, 1981) to find the best segment of similarity between two sequences.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a reference amino acid sequence of a Nodal or Lefty polypeptide is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid of the Nodal or Lefty polypeptide. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequence shown in Figures 1A and B (SEQ ID NO:2), the amino acid sequence shown in

Figures 2A and B (SEQ ID NO:4), the amino acid sequence encoded by deposited cDNA clones HTLFA20, HNGEF08, and HUKEJ46, or fragments thereof, can be determined conventionally using known computer programs such the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711). When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such that the percentage of identity is calculated over the full length of the reference amino acid sequence and that gaps in homology of up to 5% of the total number of amino acid residues in the reference sequence are allowed.

In a specific embodiment, the identity between a reference (query) sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, is determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. 6:237-245 (1990)). Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter. According to this embodiment, if the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction is made to the results to take into consideration the fact that the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity. For subject sequences truncated at the N- and C-termini, relative to the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned

with a corresponding subject residue, as a percent of the total bases of the query sequence. A determination of whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of this embodiment. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence. For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are made for the purposes of this embodiment.

The invention also encompasses fusion proteins in which the full-length Nodal or Lefty polypeptide or fragment, variant, derivative, or analog thereof is fused to an unrelated protein. These fusion proteins can be routinely designed on the basis of the Nodal or Lefty nucleotide and polypeptide sequences disclosed herein. For example, as one of skill in the art will appreciate, Nodal and/or Lefty polypeptides and fragments (including epitope-bearing fragments) thereof described herein can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric (fusion) polypeptides. These fusion proteins facilitate purification and show an increased half-life *in vivo*. This has been shown, e.g., for chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins (EP A 394,827; Traunecker, *et aL, Nature* **331**:84-86 (1988)). Fusion proteins that have a disulfide-linked dimeric structure due to the IgG part can also be more efficient in binding and neutralizing other molecules than the monomeric Nodal or Lefty proteins or protein fragments alone (Fountoulakis, *et al., J. Biochem.* **270**:3958-3964 (1995)). Examples of Nodal and Lefty fusion proteins that are encompassed by the invention include, but are not limited to, fusion of the Nodal or Lefty polypeptide sequences to any amino acid sequence that allows the fusion proteins to be displayed on the cell surface (e.g. the IgG Fc domain); or fusions to an enzyme, fluorescent protein, or luminescent protein which provides a marker function.

Antibodies

Nodal or Lefty polypeptide-specific antibodies for use in the present invention can be raised against the intact Nodal or Lefty protein or an antigenic polypeptide fragment thereof, which may be presented together with a carrier

protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier.

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')2 fragments) which are capable of specifically binding to Nodal or Lefty protein. Fab and F(ab')2 fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding of an intact antibody (Wahl, *et al., J. Nucl. Med.* **24:**316-325 (1983)). Thus, these fragments are preferred.

The antibodies of the present invention may be prepared by any of a variety of methods. For example, cells expressing the Nodal or Lefty protein or an antigenic fragment thereof can be administered to an animal in order to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of Nodal and Lefty protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or Nodal or Lefty protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology (Kohler, *et al., Nature* **256:**495 (1975); Kohler, *et al., Eur. J. Immunol.* **6:**511 (1976); Kohler, *et al., Eur. J. Immunol.* **6:**292 (1976); Hammerling, *et al., in*: *Monoclonal Antibodies and T-Cell Hybridomas*, Elsevier, N.Y., (1981) pp. 563-681)). In general, such procedures involve immunizing an animal (preferably a mouse) with a Nodal or Lefty protein antigen or, more preferably, with a Nodal or Lefty protein-expressing cell. Suitable cells can be recognized by their capacity to bind anti-Nodal or anti-Lefty protein antibody. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine

serum (inactivated at about 56° C), and supplemented with about 10 μg/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 μg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the American Type Culture Collection, Rockville, Maryland. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands and colleagues (*Gastroenterology* **80**:225-232 (1981)). The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the Nodal or Lefty protein antigen.

Alternatively, additional antibodies capable of binding to the Nodal or Lefty protein antigens may be produced in a two-step procedure through the use of anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and that, therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, Nodal or Lefty protein-specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the Nodal or Lefty protein-specific antibody can be blocked by the Nodal or Lefty protein antigen. Such antibodies comprise anti-idiotypic antibodies to the Nodal or Lefty protein-specific antibodies and can be used to immunize an animal to induce formation of further Nodal or Lefty protein-specific antibodies.

It will be appreciated that Fab and F(ab')2 and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage,

using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')2 fragments). Alternatively, Nodal or Lefty protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For *in vivo* use of anti-Nodal and anti-Lefty in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art (Morrison, *Science* **229**:1202 (1985); Oi, *et al.*, *BioTechniques* **4**:214 (1986); Cabilly, *et al.*, U.S. Patent No. 4,816,567; Taniguchi, *et al.*, EP 171496; Morrison, *et al.*, EP 173494; Neuberger, *et al.*, WO 8601533; Robinson, *et al.*, WO 8702671; Boulianne, *et al.*, *Nature* **312**:643 (1984); Neuberger, *et al.*, *Nature* **314**:268 (1985).

Cellular Growth and Differentiation-Related Disorders

Diagnosis

The present inventors have discovered that Nodal is expressed in neutrophils and testes. In addition, the present inventors have discovered that Lefty is expressed in uterine cancer, colon cancer, apoptotic T-cells, fetal heart, Wilm's Tumor tissue, frontal lobe of the brain from a patient with dementia, neutrophils, salivary gland, small intestine, 7, 8, and 12 week old human embryos, frontal cortex and hypothalamus from a patient with schizophrenia, brain from a patient with Alzheimer's Disease, adipose tissue, brown fat, TNF- and LPS-induced and uninduced bone marrow stroma, activated monocytes and macrophages, rhabdomyosarcoma, cycloheximide-treated Raji cells, breast lymph nodes, hemangiopericytoma, testes, fetal epithelium (skin), and IL-5-induced eosinophils.. For a number of cell growth and differentiation-related disorders, substantially altered (increased or decreased) levels of Nodal or Lefty gene

expression can be detected in affected tissues, cells, or bodily fluids (e.g., sera, plasma, urine, synovial fluid or spinal fluid) taken from an individual having such a disorder, relative to a "standard" Nodal or Lefty gene expression level, that is, the Nodal and Lefty expression level in affected tissues or bodily fluids from an individual not having the cell growth and differentiation disorder. Thus, the invention provides a diagnostic method useful during diagnosis of a cell growth and differentiation disorder, which involves measuring the expression level of the gene encoding the Nodal or Lefty proteins in affected tissues, cells, or body fluids from an individual and comparing the measured gene expression level with a standard Nodal or Lefty gene expression level, whereby an increase or decrease in the gene expression level compared to the standard is indicative of a cell growth and differentiation disorder.

In particular, it is believed that certain tissues in mammals with cancer of the immune or reproductive systems express significantly reduced levels of the Nodal or Lefty proteins and mRNA encoding the Nodal or Lefty proteins when compared to corresponding "standard" levels. Further, it is believed that enhanced levels of the Nodal or Lefty proteins can be detected in certain body fluids (e.g., sera, plasma, urine, and spinal fluid) from mammals with such a cancer when compared to sera from mammals of the same species not having the cancer.

Thus, the invention provides a diagnostic method useful during diagnosis of a cellular growth and differentiation disorder, including cancers, which involves measuring the expression level of the genes encoding the Nodal and Lefty proteins in tissues, cells, or body fluids from an individual and comparing the measured gene expression levels with standard Nodal and Lefty gene expression levels, whereby an increase or decrease in the gene expression level compared to the standard is indicative of a cell growth and differentiation disorder.

Where a diagnosis of a disorder in the regulation of cell growth and differentiation, including diagnosis of a tumor, has already been made according to conventional methods, the present invention is useful as a prognostic indicator, whereby patients exhibiting depressed Nodal or Lefty gene expression will experience a worse clinical outcome relative to patients expressing the gene at a level nearer the standard level.

By "assaying the expression level of the genes encoding the Nodal and Lefty polypeptides" is intended qualitatively or quantitatively measuring or estimating the level of the Nodal and Lefty polypeptides or the level of the mRNA encoding the Nodal and Lefty polypeptides in a first biological sample either directly (e.g., by determining or estimating absolute protein level or mRNA level) or relatively (e.g., by comparing to the Nodal and Lefty polypeptides levels or mRNA level in a second biological sample). Preferably, the Nodal and Lefty polypeptides levels or mRNA levels in the first biological sample is measured or estimated and compared to a standard Nodal and Lefty polypeptide level or mRNA level, the standard being taken from a second biological sample obtained from an individual not having the disorder or being determined by averaging levels from a population of individuals not having a disorder of cellular growth and differentiation. As will be appreciated in the art, once standard Nodal and Lefty polypeptides levels or mRNA levels are known, they can be used repeatedly as a standard for comparison.

By "biological sample" is intended any biological sample obtained from an individual, body fluid, cell line, tissue culture, or other source which contains Nodal and Lefty protein or mRNA. As indicated, biological samples include body fluids (such as sera, plasma, urine, synovial fluid and spinal fluid) which contain free active forms of Nodal or Lefty protein, tissues exhibiting the effects of abnormally regulated cell growth or differentiation, and other tissue sources found to express complete, mature, or active forms of the Nodal or Lefty proteins

or a Nodal or Lefty receptor. Methods for obtaining tissue biopsies and body fluids from mammals are well known in the art. Where the biological sample is to include mRNA, a tissue biopsy is the preferred source.

The present invention is useful for diagnosis or treatment of various cell growth and differentiation-related disorders in mammals, preferably humans. Such disorders include tumors, cancers, interstitial lung disease, and any disregulation of the growth and differentiation patterns of cell function including, but not limited to, autoimmunity, arthritis, leukemias, lymphomas, immunosuppression, immunity, humoral immunity, inflammatory bowel disease, myelosuppression, and the like.

Total cellular RNA can be isolated from a biological sample using any suitable technique such as the single-step guanidinium-thiocyanate-phenol-chloroform method described by Chomczynski and Sacchi (*Anal. Biochem.* **162**:156-159 (1987)). Levels of mRNA encoding the Nodal and Lefty polypeptides are then assayed using any appropriate method. These include Northern blot analysis, S1 nuclease mapping, the polymerase chain reaction (PCR), reverse transcription in combination with the polymerase chain reaction (RT-PCR), and reverse transcription in combination with the ligase chain reaction (RT-LCR).

Assaying Nodal and Lefty polypeptides levels in a biological sample can occur using antibody-based techniques. For example, Nodal and Lefty protein expression in tissues can be studied with classical immunohistological methods (Jalkanen, M., *et al., J. Cell. Biol.* **101**:976-985 (1985); Jalkanen, M., *et al., J. Cell. Biol.* **105**:3087-3096 (1987)). Other antibody-based methods useful for detecting Nodal and Lefty polypeptides gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as

iodine (^{125}I, ^{121}I), carbon (^{14}C), sulfur (^{35}S), tritium (^{3}H), indium (^{112}In), and technetium (^{99m}Tc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying Nodal and Lefty protein levels in a biological sample obtained from an individual, Nodal and Lefty polypeptides can also be detected *in vivo* by imaging. Antibody labels or markers for *in vivo* imaging of Nodal or Lefty protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A Nodal or Lefty polypeptide-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, ^{131}I, ^{112}In, ^{99m}Tc), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously or intraperitoneally) into the mammal to be examined for immune system disorder. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of ^{99m}Tc. The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain Nodal and Lefty protein. *in vivo* tumor imaging is described by Burchiel and coworkers (Chapter 13 in *Tumor Imaging: The Radiochemical Detection of Cancer*, Burchiel, S. W. and Rhodes, B. A., eds., Masson Publishing Inc. (1982)).

Treatment

As noted above, Nodal and Lefty polynucleotides and polypeptides are useful for diagnosis of conditions involving abnormally high or low expression of Nodal and Lefty activities. Given the cells and tissues where Nodal and Lefty are expressed as well as the activities modulated by Nodal and Lefty, it is readily apparent that a substantially altered (increased or decreased) level of expression of Nodal and Lefty in an individual compared to the standard or "normal" level produces pathological conditions related to the bodily system(s) in which Nodal and Lefty are expressed and/or are active.

It will also be appreciated by one of ordinary skill that, since the Nodal and Lefty proteins of the invention are members of the TGF-β superfamily the active domains of the proteins may be released in soluble form from the cells which express the Nodal and Lefty by proteolytic cleavage. Therefore, when Nodal or Lefty active domain is added from an exogenous source to cells, tissues or the body of an individual, the protein will exert its physiological activities on its target cells of that individual.

Therefore, it will be appreciated that conditions caused by a decrease in the standard or normal level of Nodal or Lefty activity in an individual, particularly disorders of cell growth and differentiation, can be treated by administration of the active form of Nodal or Lefty polypeptides. Thus, the invention also provides a method of treatment of an individual in need of an increased level of Nodal or Lefty activity comprising administering to such an individual a pharmaceutical composition comprising an amount of an isolated Nodal or Lefty polypeptide of the invention, particularly the active form of the Nodal and Lefty protein of the invention, effective to increase the Nodal and Lefty activity level in such an individual.

Since Nodal and Lefty inhibit endothelial cell function, compositions (e.g., polynucleotides, polypeptides, and fragments variants, derivatives and analogs thereof, and antibodies thereto, and angonists and antagonists thereto) corresponding to these genes may be used as anti-inflammatories. Nodal and Lefty compositions may also be employed to inhibit T-cell proliferation by the inhibition of IL-2 biosynthesis for the treatment of T-cell mediated auto-immune diseases and lymphocytic leukemias. In addition, compositions corresponding to Nodal and Lefty regulate T_{H1}/T_{H2} cytokine production. Further, Nodal and Lefty compositions may also be administered to treat or prevent inflammation, allergy, and infectious diseases or as an adjuvant for immunotherapy of tumors. Nodal and Lefty compositions may also be employed to stimulate wound healing. In this same manner, Nodal and Lefty compounds may also be employed to regulate hematopoiesis, by regulating the activation and differentiation of various hematopoietic progenitor cells, such as for example, to stimulate erythropoiesis or to stimulate the release of mature leukocytes from the bone marrow following chemotherapy, i.e., in stem cell mobilization.

Since Nodal is essential for mesoderm formation and subsequent organization of axial structures in early mouse development, the human Nodal homologue of the present invention is also likely involved developmental processes such as the correct formation of various structures or in one or more post-developmental capacities including sexual development, pituitary hormone production, and the creation of bone and cartilage, as are many of the other members of the TGF-β superfamily. Accordingly, the invention encompasses the use of Nodal compositions to regulate these processes, such as, for example, in stimulating bone and/or cartilage formation, and stimulating the production of pituitary hormone.

Since murine Lefty is important in left/right handedness of the developing organism. The homology between murine Lefty and the novel human Lefty

homologue of the present invention indicates that the novel human Lefty homologue of the present invention may also be involved in correct formation of various structures with respect to the rest of the developing organism or Lefty may also be involved in one or more post-developmental capacities including sexual development, pituitary hormone production, and the creation of bone and cartilage, as are many of the other members of the TGF-β superfamily. Accordingly, the invention encompasses the use of Nodal compositions to regulate these processes, such as, for example, in stimulating bone and/or cartilage formation, and stimulating the production of hormones in the pituitary.

Nodal and Lefty compounds may also be administered regulate or modulate cell growth and differentiation which is not necessarily associated with endogenously high or low levels of Nodal and/or Lefty. For example, Nodal and Lefty polypeptides of the present invention are useful for enhancing or enriching the growth and/or differentiation of specific cell populations, e.g., embryonic cells or stem cells.

Formulations and Administration

The Nodal and/or Lefty polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with Nodal and/or Lefty polypeptide alone), the site of delivery of the Nodal and/or Lefty polypeptide composition, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" of Nodal and/or Lefty polypeptide for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of Nodal and/or Lefty polypeptide administered parenterally per dose will be in the range of about 1 μg/kg/day to 10 mg/kg/day of patient body weight, although, as

noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the Nodal and/or Lefty polypeptide is typically administered at a dose rate of about 1 μg/kg/hour to about 50 μg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the Nodal and Lefty proteins of the invention may be administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally, topically (as by powders, ointments, drops or transdermal patch), bucally, or as an oral or nasal spray. By "pharmaceutically acceptable carrier" is meant a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The Nodal and Lefty polypeptides are also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or mirocapsules. Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U., *et al.*, *Biopolymers* **22**:547-556 (1983)), poly (2-hydroxyethyl methacrylate; Langer, R., *et al.*, *J. Biomed. Mater. Res.* **15**:167-277 (1981), and Langer, R., *Chem. Tech.* **12**:98-105 (1982)), ethylene vinyl acetate (Langer, R., *et al.*, *Id.*) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release Nodal and Lefty polypeptide compositions also include

liposomally entrapped Nodal and Lefty polypeptides. Liposomes containing Nodal and Lefty polypeptides are prepared by methods known in the art (DE 3,218,121; Epstein, *et al., Proc. Natl. Acad. Sci. (USA)* **82**:3688-3692 (1985); Hwang, *et al., Proc. Natl. Acad. Sci. (USA)* 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324). Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal Nodal and Lefty polypeptide therapy.

For parenteral administration, in one embodiment, the Nodal and/or Lefty polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the Nodal and Lefty polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include

buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

Another embodiment of the invention provides pharmaceutical compositions which contain a therapeutically effective amount of human Nodal and/or Lefty polypeptide, in a pharmaceutically acceptable vehicle or carrier. These compositions of the invention may be useful in the therapeutic modulation or diagnosis of bone, cartilage, or other connective cell or tissue growth and/or differentiation. These compositions may be used to treat such conditions as osteoarthritis, osteoporosis, and other abnormalities of bone, cartilage, muscle, tendon, ligament and/or other connective tissues and/or organs such as liver, lung, cardiac, pancreas, kidney, and other tissues. These compositions may also be useful in the growth and/or formation of cartilage, tendon, ligament, meniscus, and other connective tissues or any combination of the above (e.g., therapeutic modulation of the tendon-to-bone attachment apparatus). These compositions may also be useful in treating periodontal disease and modulating wound healing and tissue repair of such tissues as epidermis, nerve, muscle, cardiac muscle, liver, lung, cardiac, pancreas, kidney, and other tissues and/or organs. Pharmaceutical compositions containing Nodal and/or Lefty of the invention may include one or more other therapeutically useful component such as BMP-1, BMP-2, BMP-3, BMP-4, BMP-5, BMP-6, and/or BMP-7 (*See*, for example, U. S. Patent Nos. 5,108,922; 5,013,649; 5,116,738; 5,106,748; 5,187,076; and 5,141,905), BMP-8

(*See*, for example, PCT publication WO91/18098), BMP-9 (*See*, for example, PCT publication WO93/00432), BMP-10 (*See*, for example, PCT publication WO94/26893), BMP-11 (*See*, for example, PCT publication WO94/26892), BMP-12 and/or BMP-13 (*See*, for example, PCT publication WO95/16035), with other growth factors including, but not limited to, BIP, one or more of the growth and differentiation factors (GDFs), VGR-2, epidermal growth factor (EGF), fibroblast growth factor (FGF), TGF-alpha, TGF-beta, activins, inhibins, and insulin-like growth factor (IGF).

The Nodal and Lefty polypeptides are typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of Nodal and Lefty polypeptide salts.

Nodal and Lefty polypeptides to be used for therapeutic administration must be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic Nodal and Lefty polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Nodal and Lefty polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous Nodal and Lefty polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized Nodal and Lefty polypeptide using bacteriostatic water-for-injection (WFI).

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical

compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Agonists and Antagonists - Assays and Molecules

The invention also provides a method of screening compounds to identify those which enhance or block the action of Nodal and Lefty on cells, such as their interactions with Nodal- or Lefty-binding molecules such as receptor molecules. An agonist is a compound which increases the natural biological functions of Nodal or Lefty or which functions in a manner similar to Nodal or Lefty, while antagonists decrease or eliminate such functions.

In another embodiment, the invention provides a method for identifying a receptor protein or other ligand-binding protein which binds specifically to a Nodal or Lefty polypeptide. For example, a cellular compartment, such as a membrane or a preparation thereof, may be prepared from a cell that expresses a molecule that binds Nodal or Lefty. The preparation is incubated with labeled Nodal or Lefty and complexes of Nodal or Lefty bound to the receptor or other binding protein are isolated and characterized according to routine methods known in the art. Alternatively, the Nodal or Lefty polypeptides may be bound to a solid support so that binding molecules solubilized from cells are bound to the column and then eluted and characterized according to routine methods.

In the assay of the invention for agonists or antagonists, a cellular compartment, such as a membrane or a preparation thereof, may be prepared from a cell that expresses a molecule that binds Nodal or Lefty, such as a molecule of a signaling or regulatory pathway modulated by Nodal or Lefty. The

preparation is incubated with labeled Nodal or Lefty in the absence or the presence of a candidate molecule which may be a Nodal or Lefty agonist or antagonist. The ability of the candidate molecule to bind the binding molecule is reflected in decreased binding of the labeled ligand. Molecules which bind gratuitously, i.e., without inducing the effects of Nodal or Lefty on binding the Nodal or Lefty binding molecule, are most likely to be good antagonists. Molecules that bind well and elicit effects that are the same as or closely related to Nodal or Lefty are agonists.

Nodal or Lefty-like effects of potential agonists and antagonists may by measured, for instance, by determining activity of a second messenger system following interaction of the candidate molecule with a cell or appropriate cell preparation, and comparing the effect with that of Nodal or Lefty or molecules that elicit the same effects as Nodal or Lefty. Second messenger systems that may be useful in this regard include but are not limited to AMP guanylate cyclase, ion channel or phosphoinositide hydrolysis second messenger systems.

Another example of an assay for Nodal and Lefty antagonists is a competitive assay that combines Nodal or Lefty and a potential antagonist with membrane-bound Nodal or Lefty receptor molecules or recombinant Nodal or Lefty receptor molecules under appropriate conditions for a competitive inhibition assay. Nodal and Lefty can be labeled, such as by radioactivity, such that the number of Nodal or Lefty molecules bound to a receptor molecule can be determined accurately to assess the effectiveness of the potential antagonist.

Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a polypeptide of the invention and thereby inhibit or extinguish its activity. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds the same sites on a binding molecule, such as a receptor

molecule, without inducing Nodal- or Lefty-induced activities, thereby preventing the action of Nodal or Lefty by excluding Nodal or Lefty from binding.

Other potential antagonists include antisense molecules. Antisense technology can be used to control gene expression through antisense DNA or RNA or through triple-helix formation. Antisense techniques are discussed in a number of studies (for example, Okano, *J. Neurochem.* **56**:560 (1991); "Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression." CRC Press, Boca Raton, FL (1988)). Triple helix formation is discussed in a number of studies, as well (for instance, Lee, *et al., Nucleic Acids Research* **6**:3073 (1979); Cooney, *et al., Science* **241**:456 (1988); Dervan, *et al., Science* **251**:1360 (1991)). The methods are based on binding of a polynucleotide to a complementary DNA or RNA. For example, the 5' coding portion of a polynucleotide that encodes the mature polypeptide of the present invention may be used to design an antisense RNA oligonucleotide of from about 10 to 40 base pairs in length. A DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription thereby preventing transcription and the production of Nodal or Lefty. The antisense RNA oligonucleotide hybridizes to the mRNA *in vivo* and blocks translation of the mRNA molecule into Nodal and Lefty polypeptide. The oligonucleotides described above can also be delivered to cells such that the antisense RNA or DNA may be expressed *in vivo* to inhibit production of Nodal or Lefty protein.

The agonists and antagonists may be employed in a composition with a pharmaceutically acceptable carrier, e.g., as described above.

The antagonists may be employed for instance to inhibit the activation of macrophages and their precursors, and of neutrophils, basophils, B lymphocytes and some T-cell subsets, e.g., activated and CD8 cytotoxic T cells and natural killer cells, in certain autoimmune and chronic inflammatory and infective diseases. Examples of autoimmune diseases include multiple sclerosis, and

insulin-dependent diabetes. The antagonists may also be employed to treat infectious diseases including silicosis, sarcoidosis, idiopathic pulmonary fibrosis by preventing the recruitment and activation of mononuclear phagocytes. They may also be employed to treat idiopathic hyper-eosinophilic syndrome by preventing eosinophil production and stimulation. Endotoxic shock may also be treated by the antagonists by preventing the stimulation of macrophages and their production of the human chemokine polypeptides of the present invention. The antagonists may also be employed to treat histamine-mediated allergic reactions and immunological disorders including late phase allergic reactions, chronic urticaria, and atopic dermatitis by inhibiting mast cell and basophil degranulation and release of histamine. IgE-mediated allergic reactions such as allergic asthma, rhinitis, and eczema may also be treated. The antagonists may also be employed to treat chronic and acute inflammation by preventing the activation of monocytes in a wound area. Antagonists may also be employed to treat rheumatoid arthritis by preventing the activation of monocytes in the synovial fluid in the joints of patients. Monocyte activation plays a significant role in the pathogenesis of both degenerative and inflammatory arthropathies. The antagonists may be employed to interfere with the deleterious cascades attributed primarily to IL-1 and TNF, which prevents the biosynthesis of other inflammatory cytokines. In this way, the antagonists may be employed to prevent inflammation. The antagonists may also be employed to treat cases of bone marrow failure, for example, aplastic anemia and myelodysplastic syndrome. Any of the above antagonists may be employed in a composition with a pharmaceutically acceptable carrier, e.g., as hereinafter described.

Gene Mapping

The nucleic acid molecules of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can

hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. Few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal location. The mapping of DNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

In certain preferred embodiments in this regard, the cDNAs herein disclosed are used to clone genomic DNAs of Nodal and Lefty protein genes. This can be accomplished using a variety of well known techniques and libraries, which generally are available commercially. The genomic DNAs then are used for *in situ* chromosome mapping using well known techniques for this purpose.

In addition, in some cases, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the cDNA. Computer analysis of the 3' untranslated region of the gene is used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Fluorescence *in situ* hybridization ("FISH") of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with probes from the cDNA as short as 50 or 60 bp (for a review of this technique, see Verma, *et al., Human Chromosomes: A Manual Of Basic Techniques*, Pergamon Press, New York (1988)).

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, on the World Wide Web (McKusick, V. *Mendelian Inheritance In Man*, available on-line through Johns Hopkins University, Welch Medical Library). The relationship between genes

and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

Next, it is necessary to determine the differences in the cDNA or genomic sequence between affected and unaffected individuals. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

Examples

Example 1(a): Expression and Purification of "His-tagged" Nodal in E. coli

The bacterial expression vector pQE9 (pD10) is used for bacterial expression in this example. (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311). pQE9 encodes ampicillin antibiotic resistance ("Ampr") and contains a bacterial origin of replication ("ori"), an IPTG inducible promoter, a ribosome binding site ("RBS"), six codons encoding histidine residues that allow affinity purification using nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin sold by QIAGEN, Inc., *supra,* and suitable single restriction enzyme cleavage sites. These elements are arranged such that an inserted DNA fragment encoding a polypeptide expresses that polypeptide with the six His residues (i.e., a "6 X His tag") covalently linked to the amino terminus of that polypeptide.

The DNA sequence encoding the desired portion of the Nodal and Lefty protein comprising the active domain of the Nodal amino acid sequence is amplified from the deposited cDNA clone using PCR oligonucleotide primers which anneal to the amino terminal sequences of the desired portion of the Nodal and Lefty protein and to sequences in the deposited construct 3' to the cDNA coding sequence. Additional nucleotides containing restriction sites to facilitate

cloning in the pQE9 vector are added to the 5' and 3' primer sequences, respectively.

For cloning the active form of the Nodal protein, the 5' primer has the sequence 5' CGC GGA TCC CAT CAC TTG CCA GAC AGA AG 3' (SEQ ID NO:9) containing the underlined *Bam* HI restriction site followed by 20 nucleotides of the amino terminal coding sequence of the mature Nodal sequence in SEQ ID NO:2. One of ordinary skill in the art would appreciate, of course, that the point in the protein coding sequence where the 5' primer begins may be varied to amplify a DNA segment encoding any desired portion of the complete Nodal protein shorter or longer than the active form of the protein. The 3' primer has the sequence 5' GTA CGC AAG CTT GCA GGC AAA TCC AGT CTC CCT CCA GGG ATG 3' (SEQ ID NO:10) containing the underlined *Hin*d III restriction site followed by 30 nucleotides complementary to the 3' end of the coding sequence of the Nodal DNA sequence in Figure 1B.

The amplified Nodal DNA fragment and the vector pQE9 are digested with *Bam* HI and *Hin*d III and the digested DNAs are then ligated together. Insertion of the Nodal DNA into the restricted pQE9 vector places the Nodal protein coding region downstream from the IPTG-inducible promoter and in-frame with an initiating AUG and the six histidine codons.

The skilled artisan appreciates that a similar approach could easily be designed and utilized to generate a pQE9-based bacterial expression construct for the expression of Lefty protein in *E. coli*. This would be done by designing PCR primers containing similar restriction endonuclease recognition sequences combined with gene-specific sequences for Lefty and proceeding as described above.

The ligation mixture is transformed into competent *E. coli* cells using standard procedures such as those described by Sambrook and colleagues

(*Molecular Cloning: a Laboratory Manual, 2nd Ed.*; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989)). *E. coli* strain M15/rep4, containing multiple copies of the plasmid pREP4, which expresses the lac repressor and confers kanamycin resistance ("Kanr"), is used in carrying out the illustrative example described herein. This strain, which is only one of many that are suitable for expressing Nodal protein, is available commercially (QIAGEN, Inc., *supra*). Transformants are identified by their ability to grow on LB plates in the presence of ampicillin and kanamycin. Plasmid DNA is isolated from resistant colonies and the identity of the cloned DNA confirmed by restriction analysis, PCR and DNA sequencing.

Clones containing the desired constructs are grown overnight ("O/N") in liquid culture in LB media supplemented with both ampicillin (100 μg/ml) and kanamycin (25 μg/ml). The O/N culture is used to inoculate a large culture, at a dilution of approximately 1:25 to 1:250. The cells are grown to an optical density at 600 nm ("OD600") of between 0.4 and 0.6. Isopropyl-β-D-thiogalactopyranoside ("IPTG") is then added to a final concentration of 1 mM to induce transcription from the lac repressor sensitive promoter, by inactivating the lacI repressor. Cells subsequently are incubated further for 3 to 4 hours. Cells then are harvested by centrifugation.

The cells are then stirred for 3-4 hours at 4°C in 6M guanidine-HCl, pH 8. The cell debris is removed by centrifugation, and the supernatant containing the Nodal protein is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist, 1995, QIAGEN, Inc., *supra*). Briefly the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10

volumes of 6 M guanidine-HCl pH 6, and finally the Nodal is eluted with 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins can be eluted by the addition of 250 mM immidazole. Immidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4° C or frozen at -80° C.

The following alternative method may be used to purify Nodal expressed in *E coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells are harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

The cells ware then lysed by passing the solution through a microfluidizer (Microfuidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 x *g* for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 x *g* centrifugation for 15 min., the pellet is discarded and the Nodal polypeptide-containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

Following high speed centrifugation (30,000 x *g*) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

To clarify the refolded Nodal polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 μm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 mm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the Nodal polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant A_{280} monitoring of

the effluent. Fractions containing the Nodal polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant Nodal polypeptide exhibits greater than 95% purity after the above refolding and purification steps. No major contaminant bands are observed from Commassie blue stained 16% SDS-PAGE gel when 5 μg of purified protein is loaded. The purified protein is also tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

Example 2: Cloning and Expression of Nodal protein in a Baculovirus Expression System

In this illustrative example, the plasmid shuttle vector pA2GP is used to insert the cloned DNA encoding the active form of the Nodal protein, lacking its naturally associated secretory signal (leader) sequence, into a baculovirus to express the active form of the Nodal protein, using a baculovirus leader and standard methods as described by Summers and colleagues (*A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures*, Texas Agricultural Experimental Station Bulletin No. 1555 (1987)). This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by the secretory signal peptide (leader) of the baculovirus gp67 protein and convenient restriction sites such as *Bam* HI, *Xba* I and *Asp* 718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak Drosophila promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by

viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate viable virus that expresses the cloned polynucleotide.

Many other baculovirus vectors could be used in place of the vector above, such as pAc373, pVL941 and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, by Luckow and colleagues (*Virology* **170**:31-39 (1989)).

The cDNA sequence encoding the mature Nodal protein in the deposited clone, lacking the AUG initiation codon and the naturally associated leader sequence shown in SEQ ID NO:2, is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' sequences of the gene. The 5' primer has the sequence 5' CAA TTG GAT CCA CTT GCC AGA CAG AGA ACT CAA CTG 3' (SEQ ID NO:11) containing the underlined *Bam* HI restriction enzyme site followed by 25 nucleotides of the sequence of the active form of the Nodal protein shown in SEQ ID NO:2, beginning with the indicated N-terminus of the active form of the Nodal protein. The 3' primer has the sequence 5' CAC TTA GGT ACC ATG TCA TCA GAG GCA CCC ACA TTC TTC 3' (SEQ ID NO:12) containing the underlined *Asp* 718 restriction site followed by 27 nucleotides complementary to the 3' coding sequence in Figure 1B.

The skilled artisan appreciates that a similar approach could easily be designed and utilized to generate a pA2GP-based baculovirus expression construct for the expression of Lefty protein by baculovirus. This would be done by designing PCR primers containing the same, or similar, restriction endonuclease recognition sequences combined with gene-specific sequences for Lefty and proceeding as described above.

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with *Bam* HI and *Asp* 718 and again is purified on a 1% agarose gel. This fragment is designated herein F1.

The plasmid is digested with the restriction enzymes *Bam* HI and *Asp* 718 and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("Geneclean" BIO 101 Inc., La Jolla, Ca.). This vector DNA is designated herein "V1".

Fragment F1 and the dephosphorylated plasmid V1 are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Statagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria are identified that contain the plasmid with the human Nodal sequences by digesting DNA from individual colonies using *Bam* HI and *Asp* 718 and then analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing. This plasmid is designated herein pA2Nodal.

Five μg of the plasmid pA2Nodal is co-transfected with 1.0 μg of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner and colleaguew (*Proc. Natl. Acad. Sci. USA* **84**:7413-7417 (1987)). One μg of BaculoGold™ virus DNA and 5 μg of the plasmid pA2Nodal are mixed in a sterile well of a microtiter plate containing 50 μl of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 μl Lipofectin plus 90 μl Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's

medium without serum. The plate is then incubated for 5 hours at 27°C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27°C for four days.

After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith (*supra*). An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10). After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200 µl of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4°C. The recombinant virus is called V-Nodal.

To verify the expression of the active form of the Nodal protein, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus V-Nodal at a multiplicity of infection ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 µCi of ^{35}S-methionine and 5 µCi ^{35}S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins in the supernatant as well as the intracellular

proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the active form of the Nodal protein.

Example 3: Cloning and Expression of Nodal in Mammalian Cells

A typical mammalian expression vector contains the promoter element, which mediates the initiation of transcription of mRNA, the protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription can be achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLVI, HIVI and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter). Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden), pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146) and pBC12MI (ATCC 67109). Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the gene can be expressed in stable cell lines that contain the gene integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful to develop cell lines that carry several hundred or even several thousand copies of the gene of interest. Another useful selection marker is the enzyme glutamine synthase (GS; Murphy, *et al., Biochem J.* **227:**277-279 (1991); Bebbington, *et al., Bio/Technology* **10:**169-175 (1992)). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

The expression vectors pC1 and pC4 contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen, *et al., Mol. Cel. Biol.* **5:**438-447 (1985)) plus a fragment of the CMV-enhancer (Boshart, *et al., Cell* **41:**521-530 (1985)). Multiple cloning sites, e.g., with the restriction enzyme cleavage sites *Bam* HI, *Xba* I and *Asp* 718, facilitate the cloning of the gene of interest. The vectors contain in addition the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene.

Example 3(a): Cloning and Expression in COS Cells

The expression plasmid, pNodalHA, is made by cloning a portion of the cDNA encoding the active form of the Nodal protein into the expression vector pcDNAI/Amp or pcDNAIII (which can be obtained from Invitrogen, Inc.). To produce a soluble, secreted form of the polypeptide, the active form of Nodal is fused to the secretory leader sequence of the human IL-6 gene.

The expression vector pcDNAI/amp contains: (1) an *E. coli* origin of replication effective for propagation in *E. coli* and other prokaryotic cells; (2) an ampicillin resistance gene for selection of plasmid-containing prokaryotic cells;

(3) an SV40 origin of replication for propagation in eukaryotic cells; (4) a CMV promoter, a polylinker, an SV40 intron; (5) several codons encoding a hemagglutinin fragment (i.e., an "HA" tag to facilitate purification) followed by a termination codon and polyadenylation signal arranged so that a cDNA can be conveniently placed under expression control of the CMV promoter and operably linked to the SV40 intron and the polyadenylation signal by means of restriction sites in the polylinker. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein described by Wilson and colleagues (*Cell* **37**:767 (1984)). The fusion of the HA tag to the target protein allows easy detection and recovery of the recombinant protein with an antibody that recognizes the HA epitope. pcDNAIII contains, in addition, the selectable neomycin marker.

A DNA fragment encoding the active form of the Nodal polypeptide is cloned into the polylinker region of the vector so that recombinant protein expression is directed by the CMV promoter. The plasmid construction strategy is as follows. The Nodal cDNA of the deposited clone is amplified using primers that contain convenient restriction sites, much as described above for construction of vectors for expression of Nodal in *E. coli*. Suitable primers include the following, which are used in this example. The 5' primer, containing the underlined *Bam* HI site, a Kozak sequence, an AUG start codon, a sequence encoding the secretory leader peptide from the human IL-6 gene, and 27 nucleotides of the 5' coding region of the complete form of the Nodal polypeptide, has the following sequence: 5' GCC GGA TCC *GCC ACC* ATG AAC TCC TTC TCC ACA AGC GCC TTC GGT CCA GTT GCC TTC TCC CTG GGG CTG CTC CTG GTG TTG CCT GCT GCC TTC CCT GCC CCA GTC ATC ACT TGC CAG ACA GAA GTC AAC TG 3' (SEQ ID NO:13). The 3' primer, containing the underlined *Xba* I and 27 of nucleotides complementary to the 3' coding sequence immediately before the stop

codon, has the following sequence: 5' GGC TCT AGA ATG TCA TCA GAG GCA CCC ACA TTC TTC 3' (SEQ ID NO:14).

The skilled artisan appreciates that a similar approach could easily be designed and utilized to generate a pcDNAI/amp-based eukaryotic expression construct for the expression of Lefty protein by COS cells. This would be done by designing PCR primers containing the same, or similar, restriction endonuclease recognition sequences combined with gene-specific sequences for Lefty and proceeding as described above.

The PCR amplified DNA fragment and the vector, pcDNAI/Amp, are digested with *Bam* HI and *Xba* I and then ligated. The ligation mixture is transformed into *E. coli* strain SURE (Stratagene Cloning Systems, La Jolla, CA 92037), and the transformed culture is plated on ampicillin media plates which then are incubated to allow growth of ampicillin resistant colonies. Plasmid DNA is isolated from resistant colonies and examined by restriction analysis or other means for the presence of the fragment encoding the active form of the Nodal polypeptide.

For expression of recombinant Nodal, COS cells are transfected with an expression vector, as described above, using DEAE-dextran, as described, for instance, by Sambrook and coworkers (*Molecular Cloning: a Laboratory Manual*, Cold Spring Laboratory Press, Cold Spring Harbor, New York (1989)). Cells are incubated under conditions for expression of Nodal and Lefty by the vector.

Expression of the Nodal-HA fusion protein is detected by radiolabeling and immunoprecipitation, using methods described in, for example Harlow and colleagues (*Antibodies: A Laboratory Manual, 2nd Ed.*; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1988)). To this end, two days after transfection, the cells are labeled by incubation in media containing ^{35}S-cysteine for 8 hours. The cells and the media are collected, and the cells are

washed and the lysed with detergent-containing RIPA buffer: 150 mM NaCl, 1% NP-40, 0.1% SDS, 1% NP-40, 0.5% DOC, 50 mM TRIS, pH 7.5, as described by Wilson and colleagues (*supra*). Proteins are precipitated from the cell lysate and from the culture media using an HA-specific monoclonal antibody. The precipitated proteins then are analyzed by SDS-PAGE and autoradiography. An expression product of the expected size is seen in the cell lysate, which is not seen in negative controls.

Example 3(b): Cloning and Expression in CHO Cells

The vector pC4 is used for the expression of the active form of the Nodal polypeptide. Plasmid pC4 is a derivative of the plasmid pSV2-dhfr (ATCC Accession No. 37146). To produce a soluble, secreted form of the polypeptide, the active form of Nodal is fused to the secretory leader sequence of the human IL-6 gene. The plasmid contains the mouse DHFR gene under control of the SV40 early promoter. Chinese hamster ovary- or other cells lacking dihydrofolate activity that are transfected with these plasmids can be selected by growing the cells in a selective medium (alpha minus MEM, Life Technologies) supplemented with the chemotherapeutic agent methotrexate. The amplification of the DHFR genes in cells resistant to methotrexate (MTX) has been well documented (see, e.g., Alt, F. W., *et al., J. Biol. Chem.* **253:**1357-1370 (1978); Hamlin, J. L. and Ma, C. *Biochem. et Biophys. Acta,* **1097:**107-143 (1990); Page, M. J. and Sydenham, M. A. *Biotechnology* **9:**64-68 (1991)). Cells grown in increasing concentrations of MTX develop resistance to the drug by overproducing the target enzyme, DHFR, as a result of amplification of the DHFR gene. If a second gene is linked to the DHFR gene, it is usually co-amplified and over-expressed. It is known in the art that this approach may be used to develop cell lines carrying more than 1,000 copies of the amplified gene(s). Subsequently, when the

methotrexate is withdrawn, cell lines are obtained which contain the amplified gene integrated into one or more chromosome(s) of the host cell.

Plasmid pC4 contains for expressing the gene of interest the strong promoter of the long terminal repeat (LTR) of the Rouse Sarcoma Virus (Cullen, *et al., Mol. Cell. Biol.* **5**:438-447 (1985)) plus a fragment isolated from the enhancer of the immediate early gene of human cytomegalovirus (CMV; Boshart, *et al., Cell* **41**:521-530 (1985)). Downstream of the promoter are the following single restriction enzyme cleavage sites that allow the integration of the genes: *Bam* HI, *Xba* I, and *Asp* 718. Behind these cloning sites the plasmid contains the 3' intron and polyadenylation site of the rat preproinsulin gene. Other high efficiency promoters can also be used for the expression, e.g., the human ß-actin promoter, the SV40 early or late promoters or the long terminal repeats from other retroviruses, e.g., HIV and HTLVI. Clontech's Tet-Off and Tet-On gene expression systems and similar systems can be used to express the Nodal polypeptide in a regulated way in mammalian cells (Gossen, M., and Bujard, H. *Proc. Natl. Acad. Sci. USA* **89**:5547-5551 (1992)). For the polyadenylation of the mRNA other signals, e.g., from the human growth hormone or globin genes can be used as well. Stable cell lines carrying a gene of interest integrated into the chromosomes can also be selected upon co-transfection with a selectable marker such as gpt, G418 or hygromycin. It is advantageous to use more than one selectable marker in the beginning, e.g., G418 plus methotrexate.

The plasmid pC4 is digested with the restriction enzymes *Bam* HI and *Asp* 718 and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel. The DNA sequence encoding the active form of the Nodal polypeptide is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' sequences of the desired portion of the gene. The 5' primer containing the

underlined *Bam* HI site, a Kozak sequence, an AUG start codon, and 26 nucleotides of the 5' coding region of the active form of the Nodal polypeptide, has the following sequence: 5' GAC TGG ATC CCA TAC TTG CCA GAC AGA AGT CAA CTG 3' (SEQ ID NO:15). The 3' primer, containing the underlined *Bam* HI and 26 of nucleotides complementary to the 3' coding sequence immediately before the stop codon as shown in Figure 1B (SEQ ID NO:1), has the following sequence: 5' CAC TTA GGT ACC ATG TCA TCA GAG GCA CCC ACA TTC TTC 3' (SEQ ID NO:16).

The skilled artisan appreciates that a similar approach could easily be designed and utilized to generate a pC4-based eukaryotic expression construct for the expression of Lefty protein by CHO cells. This would be done by designing PCR primers containing the same, or similar, restriction endonuclease recognition sequences combined with gene-specific sequences for Lefty and proceeding as described above.

The amplified fragment is digested with the endonucleases *Bam* HI and *Asp* 718 and then purified again on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC4 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene are used for transfection. Five μg of the expression plasmid pC4 is cotransfected with 0.5 μg of the plasmid pSVneo using lipofectin (Felgner, *et al., supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner,

Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of metothrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 μM, 2 μM, 5 μM, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100-200 μM. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

Example 4: Tissue distribution of Nodal and Lefty mRNA expression

Northern blot analysis is carried out to examine Nodal and Lefty gene expression in human tissues, using methods described by, among others, Sambrook and colleagues (*supra*). A cDNA probe containing the entire nucleotide sequence of the Nodal and/or Lefty proteins (SEQ ID NO:1) is labeled with ^{32}P using the *redi*prime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using a NucTrap column (Stratagene, La Jolla, CA), according to manufacturer's protocol. The purified labeled probe is then used to examine various human tissues for Nodal and Lefty mRNA.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) are obtained from Clontech and are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C overnight, and films developed according to standard procedures.

Using a protocol such as this expression of the Nodal mRNA was detected in fetal brain, but not in most adult tissues. Furthermore, Lefty mRNA was detected in pancreas, ovary, and colon, to a lesser extent in placenta and heart, and very weakly in testes.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

The entire disclosure of all publications (including patents, patent applications, journal articles, laboratory manuals, books, or other documents) cited herein are hereby incorporated by reference.

Further, the Sequence Listing submitted herewith, and the Sequence Listing submitted with U. S. Provisional Application Serial No. 60/056,565, filed on August 21, 1997 (to which the present application claims benefit of the filing date under 35 U.S.C. § 119(e)), in both computer and paper forms are hereby incorporated by reference in their entireties.

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 4, line 6.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection

Address of depositary institution (*including postal code and country*)
10801 University Boulevard
Manassas, Virginia 20110-2209
United States of America

Date of deposit	Accession Number
June 5, 1997	209092

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications, e.g., "Accession Number of Deposit"*)

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer: Virginia L Lilly	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 4, line 8.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection

Address of depositary institution (*including postal code and country*)
10801 University Boulevard
Manassas, Virginia 20110-2209
United States of America

Date of deposit	Accession Number
July 2, 1997	209135

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications, e.g., "Accession Number of Deposit"*)

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer: Virginia L Lilly	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 4, line 22.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution
American Type Culture Collection

Address of depositary institution (*including postal code and country*)
10801 University Boulevard
Manassas, Virginia 20110-2209
United States of America

Date of deposit	Accession Number
June 5, 1997	209091

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications, e.g., "Accession Number of Deposit"*)

For receiving Office use only	For International Bureau use only
☑ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer: Virginia L. Lilly	Authorized officer

What Is Claimed Is:

1. An isolated nucleic acid molecule nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a nucleotide sequence encoding the Nodal polypeptide having the complete amino acid sequence in SEQ ID NO:2 (i.e., positions 1 to 283 of SEQ ID NO:2);

(b) a nucleotide sequence encoding the predicted active Nodal polypeptide having the amino acid sequence at positions 173 to 283 of SEQ ID NO:2;

(c) a nucleotide sequence encoding the Nodal polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 or 209135;

(d) a nucleotide sequence encoding the active domain of the Nodal polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 or 209135;

(e) a nucleotide sequence encoding the Lefty polypeptide having the complete amino acid sequence in SEQ ID NO:4 (i.e., positions -18 to 348 of SEQ ID NO:4);

(f) a nucleotide sequence encoding the Lefty polypeptide having the complete amino acid sequence in SEQ ID NO:4 excepting the N-terminal methionine (i.e., positions -17 to 348 of SEQ ID NO:4);

(g) a nucleotide sequence encoding the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 60 to 348 of SEQ ID NO:4;

(h) a nucleotide sequence encoding the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 118 to 348 of SEQ ID NO:4;

(i) a nucleotide sequence encoding the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 125 to 348 of SEQ ID NO:4;

(j) a nucleotide sequence encoding the Lefty polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091;

(k) a nucleotide sequence encoding the Lefty polypeptide having the complete amino acid sequence excepting the N-terminal methionine encoded by the cDNA clone contained in ATCC Deposit No. 209091;

(l) a nucleotide sequence encoding the active domain of the Lefty polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091; and,

(m) a nucleotide sequence complementary to any of the nucleotide sequences in (a) through (l) above.

2. The nucleic acid molecule of claim 1 wherein said polynucleotide has the complete nucleotide sequence in Figures 1A and 1B (SEQ ID NO:1) or in Figures 2A and 2B (SEQ ID NO:3).

3. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence in Figures 1A and 1B (SEQ ID NO:1) encoding the Nodal polypeptide having the amino acid sequence in positions 1 to 283 of SEQ ID NO:2.

4. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence in Figures 2A and 2B (SEQ ID NO:3) encoding the Lefty polypeptide having the amino acid sequence in positions -18 to 348 of SEQ ID NO:4.

5. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence in Figures 1A and 1B (SEQ ID NO:1) encoding the Nodal polypeptide having the amino acid sequence in positions 2 to 283 of SEQ ID NO:2.

6. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence in Figures 2A and 2B (SEQ ID NO:3) encoding the Lefty polypeptide having the amino acid sequence in positions -17 to 348 of SEQ ID NO:4.

7. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence in Figures 1A and 1B (SEQ ID NO:1) encoding the active form of the Nodal polypeptide having the amino acid sequence from about 173 to about 283 in SEQ ID NO:2.

8. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence in Figures 2A and 2B (SEQ ID NO:3) encoding the mature form of the Lefty polypeptide having the amino acid sequence from about 1 to about 348 in SEQ ID NO:4.

9. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence in Figure 2A and 2B (SEQ ID NO:3) encoding the

active form of the Lefty polypeptide having the amino acid sequence from about 60 to about 348 in SEQ ID NO:4.

10. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence in Figures 2A and 2B (SEQ ID NO:3) encoding the active form of the Lefty polypeptide having the amino acid sequence from about 118 to about 348 in SEQ ID NO:4.

11. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence in Figures 2A and 2B (SEQ ID NO:3) encoding the active form of the Lefty polypeptide having the amino acid sequence from about 125 to about 348 in SEQ ID NO:4.

12. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a nucleotide sequence encoding a polypeptide comprising the amino acid sequence of residues n-283 of SEQ ID NO:2, where n is an integer in the range of 173-183;

(b) a nucleotide sequence encoding a polypeptide comprising the amino acid sequence of residues 1-m of SEQ ID NO:2, where m is an integer in the range of 249-283;

(c) a nucleotide sequence encoding a polypeptide having the amino acid sequence consisting of residues n-m of SEQ ID NO:2, where n and m are integers as defined respectively in (a) and (b) above;

(d) a nucleotide sequence encoding a polypeptide consisting of a portion of the complete Nodal amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 or 209135 wherein said portion excludes

from 1 to about 182 amino acids from the amino terminus of said complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 or 209135;

(e) a nucleotide sequence encoding a polypeptide consisting of a portion of the complete Nodal amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 or 209135 wherein said portion excludes from 1 to about 34 amino acids from the carboxy terminus of said complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 or 209135; and

(f) a nucleotide sequence encoding a polypeptide consisting of a portion of the complete Nodal amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 or 209135 wherein said portion include a combination of any of the amino terminal and carboxy terminal deletions in (d) and (e), above.

13. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a nucleotide sequence encoding a polypeptide comprising the amino acid sequence of residues n-348 of SEQ ID NO:4, where n is an integer in the range of 1-60;

(b) a nucleotide sequence encoding a polypeptide comprising the amino acid sequence of residues n-348 of SEQ ID NO:4, where n is an integer in the range of 1-118;

(c) a nucleotide sequence encoding a polypeptide comprising the amino acid sequence of residues n-348 of SEQ ID NO:4, where n is an integer in the range of 1-125;

(d) a nucleotide sequence encoding a polypeptide comprising the amino acid sequence of residues 1-m of SEQ ID NO:4, where m is an integer in the range of 335-348;

(e) a nucleotide sequence encoding a polypeptide having the amino acid sequence consisting of residues n-m of SEQ ID NO:4, where n and m are integers as defined respectively in (a) through (d) above;

(f) a nucleotide sequence encoding a polypeptide consisting of a portion of the complete Lefty amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091 wherein said portion excludes from 1 to about 78 amino acids from the amino terminus of said complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091;

(g) a nucleotide sequence encoding a polypeptide consisting of a portion of the complete Lefty amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091 wherein said portion excludes from 1 to about 136 amino acids from the amino terminus of said complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091;

(h) a nucleotide sequence encoding a polypeptide consisting of a portion of the complete Lefty amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091 wherein said portion excludes from 1 to about 143 amino acids from the amino terminus of said complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091;

(i) a nucleotide sequence encoding a polypeptide consisting of a portion of the complete Lefty amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091 wherein said portion excludes from 1 to about 13 amino acids from the carboxy terminus of said complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091; and

(f) a nucleotide sequence encoding a polypeptide consisting of a portion of the complete Lefty amino acid sequence encoded by the cDNA clone contained

in ATCC Deposit No. 209091 wherein said portion include a combination of any of the amino terminal and carboxy terminal deletions in (f) through (i), above.

14. The nucleic acid molecule of claim 1 wherein said polynucleotide has the complete nucleotide sequence of the cDNA clone contained in ATCC Deposit No. 209092, 209135 or 209091.

15. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence encoding the Nodal or Lefty polypeptides having the complete amino acid sequence excepting the N-terminal methionine encoded by the cDNA clones contained in ATCC Deposit No. 209092, 209135 or 209091.

16. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence encoding the mature form of the Lefty polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091.

17. The nucleic acid molecule of claim 1 wherein said polynucleotide has the nucleotide sequence encoding the active forms of the Nodal or Lefty polypeptides having the amino acid sequence encoded by the cDNA clones contained in ATCC Deposit No. 209092, 209135 or 209091.

18. An isolated nucleic acid molecule comprising a polynucleotide which hybridizes under stringent hybridization conditions to a polynucleotide having a nucleotide sequence identical to a nucleotide sequence in (a) through (m) of claim 1 wherein said polynucleotide which hybridizes does not hybridize under stringent hybridization conditions to a polynucleotide having a nucleotide sequence consisting of only A residues or of only T residues.

19. An isolated nucleic acid molecule comprising a polynucleotide which encodes the amino acid sequence of an epitope-bearing portion of a Nodal or Lefty polypeptide having an amino acid sequence in (a)through (m) of claim 1.

20. The isolated nucleic acid molecule of claim 19, which encodes an epitope-bearing portion of a Nodal polypeptide wherein the amino acid sequence of said portion is selected from the group of sequences in SEQ ID NO:2 consisting of: about Lys-54 to about Asp-62, from about Val-91 to about Leu-99, from about Lys-100 to about Gln-108, from about Cys-116 to about Pro-124, from about Gln-140 to about Leu-148, from about Trp-156 to about Ser-164, from about Arg-170, to about Gln-181, from about Cys-212 to about Phe-224, from about Tyr-239, to about Thr-247, from about Pro-251, to about Met-259, and from about Asp-263, to about His-271.

21. The isolated nucleic acid molecule of claim 19, which encodes an epitope-bearing portion of a Nodal polypeptide wherein the amino acid sequence of said portion is selected from the group of sequences in SEQ ID NO:4 consisting of: about Asp-71 to about Ser-79, from about Arg-106 to about Val-114, from about Leu-136 to about Arg-144, from about Asp-154 to about Asp-164, from about His-171 to about Asp-179, from about Gln-189 to about Leu-197, from about Pro-227 to about Glu-236, from about Gly-246 to about Glu-254, from about Pro-256 to about Gln-266, from about Cys-297 to about Ala-305, from about Ile-317 to about Pro-325, from about Ile-330 to about Val-340, and from about Val-348 to about Pro-366.

22. A recombinant vector that contains the polynucleotide of claim 1.

23. A recombinant vector that contains the polynucleotide of claim 1 operably associated with a regulatory sequence that controls gene expression.

24. A genetically engineered host cell that contains the polynucleotide of claim 1.

25. A genetically engineered host cell that contains the polynucleotide of claim 1 operatively associated with a regulatory sequence that controls gene expression.

26. A method for producing a Nodal or Lefty polypeptide, comprising; (a) culturing the genetically engineered host cell of claim 25 under conditions suitable to produce the polypeptide; and

(b) recovering said polypeptide.

27. An isolated Nodal and Lefty polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:

(a) the amino acid sequence of the full-length Nodal polypeptide having the complete amino acid sequence shown in SEQ ID NO:2 (i.e., positions 1 to 283 of SEQ ID NO:2);

(b) the amino acid sequence of the predicted active Nodal polypeptide having the amino acid sequence at positions 173 to 283 of SEQ ID NO:2;

(c) the amino acid sequence of the Nodal polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 or 209135;

(d) the amino acid sequence of the active domain of the Nodal polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209092 or 209135;

(e) the amino acid sequence of the Lefty polypeptide having the complete amino acid sequence in SEQ ID NO:4 (i.e., positions -18 to 348 of SEQ ID NO:4);

(f) the amino acid sequence of the Lefty polypeptide having the complete amino acid sequence in SEQ ID NO:4 excepting the N-terminal methionine (i.e., positions -17 to 348 of SEQ ID NO:4);

(g) the amino acid sequence of the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 60 to 348 of SEQ ID NO:4;

(h) the amino acid sequence of the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 118 to 348 of SEQ ID NO:4;

(i) the amino acid sequence of the predicted active domain of the Lefty polypeptide having the amino acid sequence at positions 125 to 348 of SEQ ID NO:4;

(j) the amino acid sequence of the Lefty polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091;

(k) the amino acid sequence of the Lefty polypeptide having the complete amino acid sequence excepting the N-terminal methionine encoded by the cDNA clone contained in ATCC Deposit No. 209091, and;

(l) the amino acid sequence of the active domain of the Lefty polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209091.

28. An isolated polypeptide comprising an epitope-bearing portion of the Nodal protein, wherein said portion is selected from the group consisting of: a polypeptide comprising amino acid residues from about Lys-54 to about Asp-62 of SEQ ID NO:2, a polypeptide comprising amino acid residues from about Val-91 to about Leu-99 of SEQ ID NO:2, a polypeptide comprising amino acid residues from about Lys-100 to about Gln-108 of SEQ ID NO:2, a polypeptide comprising amino acid residues from about Cys-116 to about Pro-124 of SEQ ID NO:2, a polypeptide comprising amino acid residues from about Gln-140 to about Leu-148 of SEQ ID NO:2, a polypeptide comprising amino acid residues from about Trp-156 to about Ser-164 of SEQ ID NO:2, a polypeptide comprising amino acid residues from about Arg-170 to about Gln-181 of SEQ ID NO:2, a polypeptide comprising amino acid residues from about Cys-212 to about Phe-224 of SEQ ID NO:2, a polypeptide comprising amino acid residues from about Tyr-239 to about Thr-247 of SEQ ID NO:2, a polypeptide comprising amino acid residues from about Pro-251 to about Met-259 of SEQ ID NO:2, and a polypeptide comprising amino acid residues from about Asp-263 to about His-271 of SEQ ID NO:2.

29. An isolated polypeptide comprising an epitope-bearing portion of the Lefty protein, wherein said portion is selected from the group consisting of: a polypeptide comprising amino acid residues from about Asp-71 to about Ser-79 of SEQ ID NO:4, a polypeptide comprising amino acid residues from about Arg-106 to about Val-114 of SEQ ID NO:4, a polypeptide comprising amino acid residues from about Leu-136 to about Arg-144 of SEQ ID NO:4, a polypeptide comprising amino acid residues from about Asp-154 to about Asp-164 of SEQ ID NO:4, a polypeptide comprising amino acid residues from about His-171 to about Asp-179 of SEQ ID NO:4, a polypeptide comprising amino acid residues from about Gln-189 to about Leu-197 of SEQ ID NO:4, a polypeptide

comprising amino acid residues from about Pro-227 to about Glu-236 of SEQ ID NO:4, a polypeptide comprising amino acid residues from about Gly-246 to about Glu-254 of SEQ ID NO:4, a polypeptide comprising amino acid residues from about Pro-256 to about Gln-266 of SEQ ID NO:4, from about Cys-297 to about Ala-305 of SEQ ID NO:4, a polypeptide comprising amino acid residues from about Ile-317 to about Pro-325 of SEQ ID NO:4, a polypeptide comprising amino acid residues from about Ile-330 to about Val-340 of SEQ ID NO:4, and a polypeptide comprising amino acid residues from about Val-348 to about Pro-366 of SEQ ID NO:4.

30. An isolated antibody that binds specifically to a Nodal and Lefty polypeptide of claim 27.

31. An isolated nucleic acid molecule comprising a polynucleotide having a sequence at least 95% identical to a sequence selected from the group consisting of:

(a) the nucleotide sequence of SEQ ID NO:7);

(b) the nucleotide sequence of SEQ ID NO:8);

(c) the nucleotide sequence of a portion of the sequence shown in Figures 1A and 1B (SEQ ID NO:1) wherein said portion comprises at least 50 contiguous nucleotides from nucleotide 1 to nucleotide 1130;

(d) the nucleotide sequence of a portion of the sequence shown in Figures 1A and 1B (SEQ ID NO:1) wherein said portion consists of nucleotides 250-1130, 500-1130, 750-1130, 1000-1130, 1-1000, 250-1000, 500-1000, 750-1000, 1-750, 250-750, 500-750, 1-500, 250-500, and 1-250 of SEQ ID NO:1;

(e) the nucleotide sequence of a portion of the sequence shown in Figures 2A and 2B (SEQ ID NO:3) wherein said portion comprises at least 50 contiguous nucleotides from nucleotide 1 to 950 and 1150 to 1688;

(f) the nucleotide sequence of a portion of the sequence shown in Figures 2A and 2B (SEQ ID NO:3) wherein said portion consists of nucleotides 250-1688, 500-1688, 750-1688, 1000-1688, 1250-1688, 1500-1688, 1-1500, 250-1500, 500-1500, 750-1500, 1000-1500, 1250-1500, 1-1250, 250-1250, 500-1250, 750-1250, 1000-1250, 1-1000, 250-1000, 500-1000, 750-1000, 1-750, 250-750, 500-750, 1-500, and 250-500 of SEQ ID NO:3; and

(g) a nucleotide sequence complementary to any of the nucleotide sequences in (a) through (f) above.

32. A method for preventing, treating, or ameliorating a medical condition which comprises administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 27.

33. A method for preventing, treating, or ameliorating a medical condition which comprises administering to a mammalian subject a therapeutically effective amount of the polynucleotide of claim 1.

34. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject related to expression or activity of Nodal or Lefty comprising:

(a) determining the presence or absence of a mutation in the polynucleotide of claim 1;

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.

35. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject related to expression or activity of Nodal or Lefty comprising:

(a) determining the presence or amount of expression of the polypeptide of claim 27 in a biological sample;

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.

36. A method of identifying compounds capable of enhancing or inhibiting a Nodal or Lefty activity comprising:

(a) contacting the polypeptide of claim 27, with a candidate compound; and

(b) assaying for activity.

Figure 1A
Nodal

```
  1  GATGTGGCAGTGGATGGGCAGAACTGGACGTTTGCTTTTGACTTCTCCTTCCTGAGCCAA   60
  1  D  V  A  V  D  G  Q  N  W  T  F  A  F  D  F  S  F  L  S  Q    20

 61  CAAGAGGATCTGGCATGGGCTGAGCTCCGGCTGCAGCTGTCCAGCCCTGTGGACCTCCCC  120
 21  Q  E  D  L  A  W  A  E  L  R  L  Q  L  S  S  P  V  D  L  P    40

121  ACTGAGGGCTCACTTGCCATTGAGATTTTCCACCAGCCAAAGCCCGACACAGAGCAGGCT  180
 41  T  E  G  S  L  A  I  E  I  F  H  Q  P  K  P  D  T  E  Q  A    60

181  TCAGACAGCTGCTTAGAGCGGTTTCAGATGGACCTATTCACTGTCACTTTGTCCCAGGTC  240
 61  S  D  S  C  L  E  R  F  Q  M  D  L  F  T  V  T  L  S  Q  V    80

241  ACCTTTTCCTTGGGCAGCATGGTTTTGGAGGTGACCAGGCCTCTCTCCAAGTGGCTGAAG  300
 81  T  F  S  L  G  S  M  V  L  E  V  T  R  P  L  S  K  W  L  K   100

301  CGCCCTGGGGCCCTGGAGAAGCAGATGTCCAGGGTAGCTGGAGAGTGCTGGCCGCGGCCC  360
101  R  P  G  A  L  E  K  Q  M  S  R  V  A  G  E  C  W  P  R  P   120

361  CCCACACCGCCTGCCACCAATGTGCTCCTTATGCTCTACTCCAACCTCTCGCAGGAGCAG  420
121  P  T  P  P  A  T  N  V  L  L  M  L  Y  S  N  L  S  Q  E  Q   140

421  AGGCAGCTGGGTGGGTCCACCTTGCTGTGGGAAGCCGAGAGCTCCTGGCGGGCCCAGGAG  480
141  R  Q  L  G  G  S  T  L  L  W  E  A  E  S  S  W  R  A  Q  E   160

481  GGACAGCTGTCCTGGGAGTGGGGCAAGAGGCACCGTCGACATCACTTGCCAGACAGAAGT  540
161  G  Q  L  S  W  E  W  G  K  R  H  R  R  H  H  L  P  D  R  S   180

541  CAACTGTGTCGGAAGGTCAAGTTCCAGGTGGACTTCAACCTGATCGGATGGGGCTCCTGG  600
181  Q  L  C  R  K  V  K  F  Q  V  D  F  N  L  I  G  W  G  S  W   200

601  ATCATCTACCCCAAGCAGTACAACGCCTATCGCTGTGAGGGCGAGTGTCCTAATCCTGTT  660
201  I  I  Y  P  K  Q  Y  N  A  Y  R  C  E  G  E  C  P  N  P  V   220

661  GGGGAGGAGTTTCATCCGACCAACCATGCATACATCCAGAGTCTGCTGAAACGTTACCAG  720
221  G  E  E  F  H  P  T  N  H  A  Y  I  Q  S  L  L  K  R  Y  Q   240

721  CCCCACCGAGTCCCTTCCACTTGTTGTGCCCCAGTGAAGACCAAGCCGCTGAGCATGCTG  780
241  P  H  R  V  P  S  T  C  C  A  P  V  K  T  K  P  L  S  M  L   260

781  TATGTGGATAATGGCAGAGTGCTCCTAGATCACCATAAAGACATGATCGTGGAAGAATGT  840
261  Y  V  D  N  G  R  V  L  L  D  H  H  K  D  M  I  V  E  E  C   280

841  GGGTGCCTCTGATGACATCCTGGAGGGAGACTGGATTTGCCTGCACTCTGGAAGGCTGGG  900
281  G  C  L  *                                                    300
```

Figure 1A (continued)
Nodal

```
901  AAACTCCTGGAAGACATGATAACCATCTAATCCAGTAAGGAGAAACAGAGAGGGGCAAAG  960

961  TTGCTCTGCCCACCAGAACTGAAGAGGAGGGGCTGCCCACTCTGTAAATGAAGGGCTCAG  1020

1021 TGGAGTCTGGCCAAGCACAGAGGCTGCTGTCAGGAAGAGGGAGGAAGAAGCCTGTGCAGG  1080

1081 GGGCTGGCTGGATGTTCTCTTTACTGAAAAGACAGTGGCAAGGAAAAGCAAAAAAAAAAAA  1140

1141 AAAAAAAAAAAAAAAA 1156
```

Figure 1B

Lefty

```
   1 GCCTTCTCAAGGGACAGCCCCACTCTGCCTCTTGCTCCTCCAGGGCAGCACCATGCAGCC   60
   1                                                   M  Q  P     3

  61 CCTGTGGCTCTGCTGGGCACTCTGGGTGTTGCCCCTGGCCAGCCCCGGGGCCGCCCTGAC  120
   4  L  W  L  C  W  A  L  W  V  L  P  L  A  S  P  G  A  A  L  T   23

 121 CGGGGAGCAGCTCCTGGGCAGCCTGCTGCGGCAGCTGCAGCTCAAAGAGGTGCCCACCCT  180
  24  G  E  Q  L  L  G  S  L  L  R  Q  L  Q  L  K  E  V  P  T  L   43

 181 GGACAGGGCCGACATGGAGGAGCTGGTCATCCCCACCCACGTGAGGGCCCAGTACGTGGC  240
  44  D  R  A  D  M  E  E  L  V  I  P  T  H  V  R  A  Q  Y  V  A   63

 241 CCTGCTGCAGCGCAGCCACGGGGACCGCTCCCGCGGAAAGAGGTTCAGCCAGAGCTTCCG  300
  64  L  L  Q  R  S  H  G  D  R  S  R  G  K  R  F  S  Q  S  F  R   83

 301 AGAGGTGGCCGGCAGGTTCCTGGCGTTGGAGGCCAGCACACACCTGCTGGTGTTCGGCAT  360
  84  E  V  A  G  R  F  L  A  L  E  A  S  T  H  L  L  V  F  G  M  103

 361 GGAGCAGCGGCTGCCGCCCAACAGCGAGCTGGTGCAGGCCGTGCTGCGGCTCTTCCAGGA  420
 104  E  Q  R  L  P  P  N  S  E  L  V  Q  A  V  L  R  L  F  Q  E  123

 421 GCCGGTCCCCAAGGCCGCGCTGCACAGGCACGGGCGGCTGTCCCCGCGCAGCGCCCGGGC  480
 124  P  V  P  K  A  A  L  H  R  H  G  R  L  S  P  R  S  A  R  A  143

 481 CCGGGTGACCGTCGAGTGGCTGCGCGTCCGCGACGACGGCTCCAACCGCACCTCCCTCAT  540
 144  R  V  T  V  E  W  L  R  V  R  D  D  G  S  N  R  T  S  L  I  163

 541 CGACTCCAGGCTGGTGTCCGTCCACGAGAGCGGCTGGAAGGCCTTCGACGTGACCGAGGC  600
 164  D  S  R  L  V  S  V  H  E  S  G  W  K  A  F  D  V  T  E  A  183

 601 CGTGAACTTCTGGCAGCAGCTGAGCCGGCCCCGGCAGCCGCTGCTGCTACAGGTGTCGGT  660
 184  V  N  F  W  Q  Q  L  S  R  P  R  Q  P  L  L  L  Q  V  S  V  203

 661 GCAGAGGGAGCATCTGGGCCCGCTGGCGTCCGGCGCCCACAAGCTGGTCCGCTTTGCCTC  720
 204  Q  R  E  H  L  G  P  L  A  S  G  A  H  K  L  V  R  F  A  S  223

 721 GCAGGGGGCGCCAGCCGGGCTTGGGGAGCCCCAGCTGGAGCTGCACACCCTGGACCTTGG  780
 224  Q  G  A  P  A  G  L  G  E  P  Q  L  E  L  H  T  L  D  L  G  243

 781 GGACTATGGAGCTCAGGGCGACTGTGACCCTGAAGCACCAATGACCGAGGGCACCCGCTG  840
 244  D  Y  G  A  Q  G  D  C  D  P  E  A  P  M  T  E  G  T  R  C  263

 841 CTGCCGCCAGGAGATGTACATTGACCTGCAGGGGATGAAGTGGGCCGAGAACTGGGTGCT  900
 264  C  R  Q  E  M  Y  I  D  L  Q  G  M  K  W  A  E  N  W  V  L  283
```

Figure 1B (continued)
Lefty

```
901  GGAGCCCCCGGGCTTCCTGGCTTATGAGTGTGTGGGCACCTGCCGGCAGCCCCCGGAGGC  960
284   E  P  P  G  F  L  A  Y  E  C  V  G  T  C  R  Q  P  P  E  A   303

961  CCTGGCCTTCAAGTGGCCGTTTCTGGGGCCTCGACAGTGCATCGCCTCGGAGACTGACTC  1020
304   L  A  F  K  W  P  F  L  G  P  R  Q  C  I  A  S  E  T  D  S   323

1021 GCTGCCCATGATCGTCAGCATCAAGGAGGGAGGCAGGACCAGGCCCCAGGTGGTCAGCCT  1080
324   L  P  M  I  V  S  I  K  E  G  G  R  T  R  P  Q  V  V  S  L   343

1081 GCCCAACATGAGGGTGCAGAAGTGCAGCTGTGCCTCGGATGGTGCGCTCGTGCCAAGGAG  1140
344   P  N  M  R  V  Q  K  C  S  C  A  S  D  G  A  L  V  P  R  R   363

1141 GCTCCAGCCATAGGCGCCTAGTGTAGCCATCGAGGGACTTGACTTGTGTGTGTTTCTGAA  1200
364   L  Q  P  *                                                    366

1201 GTGTTCGAGGGTACCAGGAGAGCTGGCGATGACTGAACTGCTGATGGACAAATGCTCTGT  1260

1261 GCTCTCTATGAGCCCTGAATTTGCTTCCTCTGACAAGTTACCTCACCTAATTTTTGCTTC  1320

1321 TCAGGAATGAGAATCTTTGGCCACTGGAGAGCCCTTGCTCAGTTTTCTCTATTCTTATTA  1380

1381 TTCACTGCACTATATTCTAAGCACTTACATGTGGAGATACTGTAACCTGAGGGCAGAAAG  1440

1441 CCCAATGTGTCATTGTTTACTTGTCCTGTCACTGGATCTGGGCTAAAGTCCTCCACCACC  1500

1501 ACTCTGGACCTAAGACCTGGGGTTAAGTGTGGGTTGTGCATCCCCAATCCAGATAATAAA  1560

1561 GACTTTGTAAAACATGAATAAAACACATTTTATTCTAAAAAAAAAAAACGGCACGAGGGGG  1620

1621 GGCCCGGTACCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTA  1680

1681 CAACGTCG  1688
```

Figure 2A
Nodal

Percent Similarity: 87.279 Percent Identity: 80.919

HNGEF08
x
muNodal

```
  1 DVAVDGQNWTFAFDFSFLSQQEDLAWAELRLQLSSPVDLPTEGSLAIEIF 50
    ||.|.||||||.||||||||:|||.||::||||.:|:|:||||.|.|:||
 66 DVDVIGQNWTFTFDFSFLSQEEDLVWADVRLQLPGPMDIPTEGPLTIDIF 115

 51 HQPKPDTEQASDSCLERFQMDLFTVTLSQVTFSLGSMVLEVTRPLSKWLK 100
    ||:|.|.|.....||||: |: |||. |||||. || |||||:|||||||
116 HQAKGDPERDPADCLERIWMETFTVIPSQVTFASGSTVLEVTKPLSKWLK 165

101 RPGALEKQMSRVAGECWPRPPTPP...ATNVLLMLYSNLSQEQRQLGGST 147
     | |||||:|. |:.||..| |||   |.. :|||||  .||||||||.|
166 DPRALEKQVSSRAEKCWHQPYTPPVPVASTNVIMLYSNRPQEQRQLGGAT 215

148 LLWEAESSWRAQEGQLSWE...WGKRHRRHHLPDRSQLCRKVKFQVDFNL 194
    ||||||||||||||||| |   ||:|:|||||||||||||:|||||||||
216 LLWEAESSWRAQEGQLSVERGGWGRRQRRHHLPDRSQLCRRVKFQVDFNL 265

195 IGWGSWIIYPKQYNAYRCEGECPNPVGEEFHPTNHAYIQSLLKRYQPHRV 244
    ||||||||||||||||||||||||||||||||||||||||||||||||||
266 IGWGSWIIYPKQYNAYRCEGECPNPVGEEFHPTNHAYIQSLLKRYQPHRV 315

245 PSTCCAPVKTKPLSMLYVDNGRVLLDHHKDMIVEECGCL 283
    |||||||||||||||||||||||:||||||||||||
316 PSTCCAPVKTKPLSMLYVDNGRVLLEHHKDMIVEECGCL 354
```

Figure 2B
Lefty

Percent Similarity: 88.525 Percent Identity: 81.967

HUKEJ46
x
muLefty

```
  1 MQPLWLCWALWVLPLASPGAALTGEQLLGSLLRQLQLKEVPTLDRADMEE 50
    |. |||||||||.|.|.|  .||||||:|||||.||||.:.|.||:||:|:
  1 MPFLWLCWALWALSLVSLREALTGEQILGSLLQQLQLDQPPVLDKADVEG 50

 51 LVIPTHVRAQYVALLQRSHGDRSRGKRFSQSFREVAGRFLALEASTHLLV 100
    :|||.|||.|||||||:||:.|||||||||.:|||||||. |.||||||
 51 MVIPSHVRTQYVALLQHSHASRSRGKRFSQNLREVAGRFLVSETSTHLLV 100

101 FGMEQRLPPNSELVQAVLRLFQEPVPKAALHRHGRLSPRSARARVTVEWL 150
    |||||||||||||||||||||||||:.||:|: ||||:|||||||:|||
101 FGMEQRLPPNSELVQAVLRLFQEPVPRTALRRQKRLSPHSARARVTIEWL 150

151 RVRDDGSNRTSLIDSRLVSVHESGWKAFDVTEAVNFWQQLSRPRQPLLLQ 200
    |.|||||||.||||||||:||||||||||||||||||||||||||||||
151 RFRDDGSNRTALIDSRLVSIHESGWKAFDVTEAVNFWQQLSRPRQPLLLQ 200

201 VSVQREHLGPLASGAHKLVRFASQGAPA..GLGEPQLELHTLDLGDYGAQ 248
    |||||||||| ..:.|||||||.||.|.  | ||||||||||| |||||
201 VSVQREHLGPGTWSSHKLVRFAAQGTPDGKGQGEPQLELHTLDLKDYGAQ 250

249 GDCDPEAPMTEGTRCCRQEMYIDLQGMKWAENWVLEPPGFLAYECVGTCR 298
    |:|||||:|||||||||||||:||||||||||:|||||||.||||||.|
251 GNCDPEAPVTEGTRCCRQEMYLDLQGMKWAENWILEPPGFLTYECVGSCL 300

299 QPPEALAFKWPFLGPRQCIASETDSLPMIVSIKEGGRTRPQVVSLPNMRV 348
    | ||.|. :|||||||||:||| .|||||||:||||||||||||||||||
301 QLPESLTSRWPFLGPRQCVASEMTSLPMIVSVKEGGRTRPQVVSLPNMRV 350

349 QKCSCASDGALVPRRLQP 366
    |.||||||||||:||||||
351 QTCSCASDGALIPRRLQP 368
```

Figure 3A
Nodal

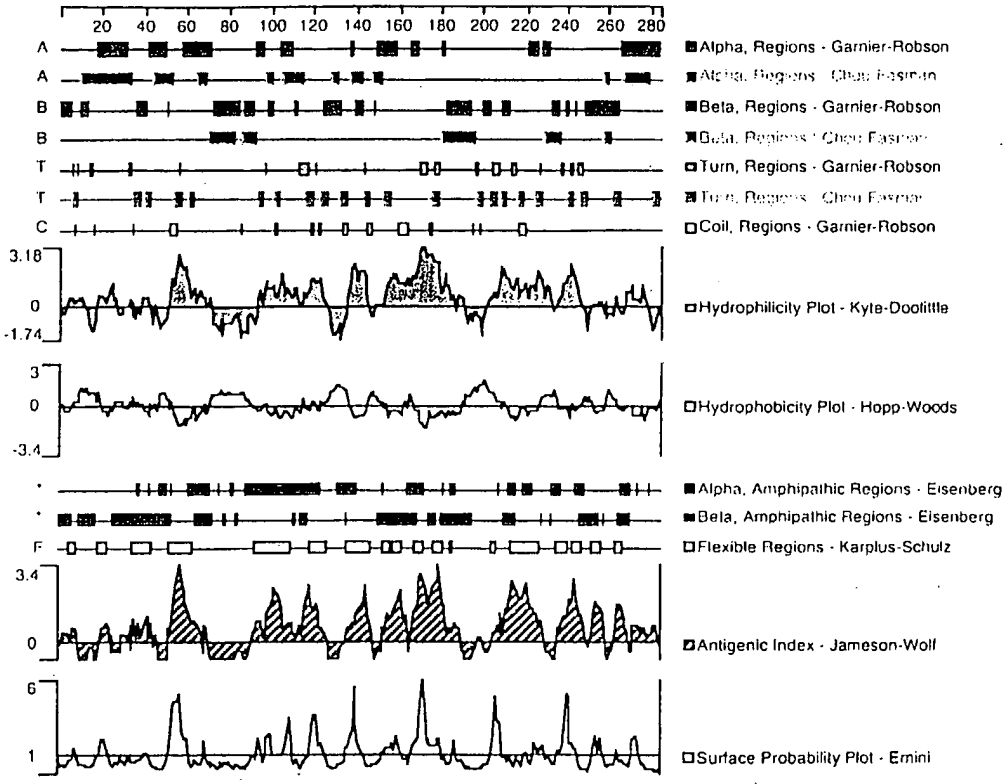
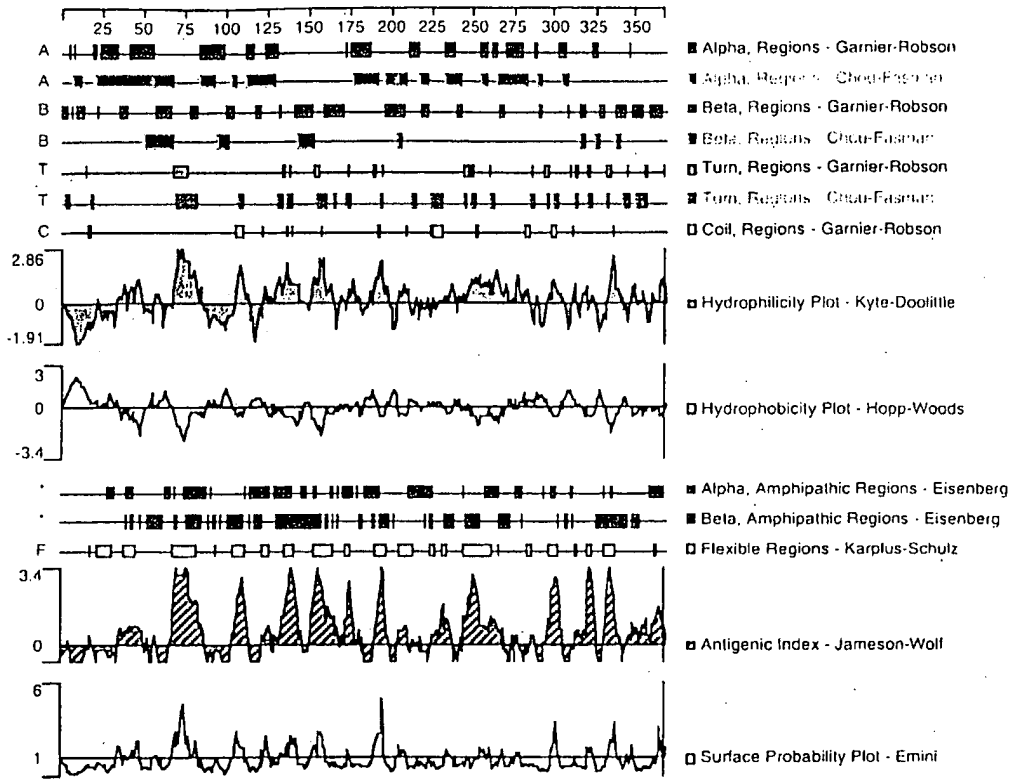


Figure 3B
Lefty



SEQUENCE LISTING

<110> Ebner, et al
Human Genome Sciences, Inc. et al.

<120> Human Nodal and Lefty Polypeptides

<130> PF380

<140> Unassigned
<141> 1998-08-20

<150> 60/056,565
<151> 1997-08-21

<160> 16

<170> PatentIn Ver. 2.0

<210> 1
<211> 1156
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(849)

<220>
<221> sig_peptide
<222> (1)..(849)

<400> 1

```
gat gtg gca gtg gat ggg cag aac tgg acg ttt gct ttt gac ttc tcc   48
Asp Val Ala Val Asp Gly Gln Asn Trp Thr Phe Ala Phe Asp Phe Ser
  1               5                  10                  15

ttc ctg agc caa caa gag gat ctg gca tgg gct gag ctc cgg ctg cag   96
Phe Leu Ser Gln Gln Glu Asp Leu Ala Trp Ala Glu Leu Arg Leu Gln
             20                  25                  30

ctg tcc agc cct gtg gac ctc ccc act gag ggc tca ctt gcc att gag   144
Leu Ser Ser Pro Val Asp Leu Pro Thr Glu Gly Ser Leu Ala Ile Glu
         35                  40                  45

att ttc cac cag cca aag ccc gac aca gag cag gct tca gac agc tgc   192
Ile Phe His Gln Pro Lys Pro Asp Thr Glu Gln Ala Ser Asp Ser Cys
     50                  55                  60

tta gag cgg ttt cag atg gac cta ttc act gtc act ttg tcc cag gtc   240
Leu Glu Arg Phe Gln Met Asp Leu Phe Thr Val Thr Leu Ser Gln Val
 65                  70                  75                  80

acc ttt tcc ttg ggc agc atg gtt ttg gag gtg acc agg cct ctc tcc   288
Thr Phe Ser Leu Gly Ser Met Val Leu Glu Val Thr Arg Pro Leu Ser
                 85                  90                  95

aag tgg ctg aag cgc cct ggg gcc ctg gag aag cag atg tcc agg gta   336
Lys Trp Leu Lys Arg Pro Gly Ala Leu Glu Lys Gln Met Ser Arg Val
            100                 105                 110

gct gga gag tgc tgg ccg cgg ccc ccc aca ccg cct gcc acc aat gtg   384
```

```
Ala Gly Glu Cys Trp Pro Arg Pro Pro Thr Pro Pro Ala Thr Asn Val
        115                 120                 125

ctc ctt atg ctc tac tcc aac ctc tcg cag gag cag agg cag ctg ggt   432
Leu Leu Met Leu Tyr Ser Asn Leu Ser Gln Glu Gln Arg Gln Leu Gly
    130                 135                 140

ggg tcc acc ttg ctg tgg gaa gcc gag agc tcc tgg cgg gcc cag gag   480
Gly Ser Thr Leu Leu Trp Glu Ala Glu Ser Ser Trp Arg Ala Gln Glu
145                 150                 155                 160

gga cag ctg tcc tgg gag tgg ggc aag agg cac cgt cga cat cac ttg   528
Gly Gln Leu Ser Trp Glu Trp Gly Lys Arg His Arg Arg His His Leu
                165                 170                 175

cca gac aga agt caa ctg tgt cgg aag gtc aag ttc cag gtg gac ttc   576
Pro Asp Arg Ser Gln Leu Cys Arg Lys Val Lys Phe Gln Val Asp Phe
            180                 185                 190

aac ctg atc gga tgg ggc tcc tgg atc atc tac ccc aag cag tac aac   624
Asn Leu Ile Gly Trp Gly Ser Trp Ile Ile Tyr Pro Lys Gln Tyr Asn
        195                 200                 205

gcc tat cgc tgt gag ggc gag tgt cct aat cct gtt ggg gag gag ttt   672
Ala Tyr Arg Cys Glu Gly Glu Cys Pro Asn Pro Val Gly Glu Glu Phe
    210                 215                 220

cat ccg acc aac cat gca tac atc cag agt ctg ctg aaa cgt tac cag   720
His Pro Thr Asn His Ala Tyr Ile Gln Ser Leu Leu Lys Arg Tyr Gln
225                 230                 235                 240

ccc cac cga gtc cct tcc act tgt tgt gcc cca gtg aag acc aag ccg   768
Pro His Arg Val Pro Ser Thr Cys Cys Ala Pro Val Lys Thr Lys Pro
                245                 250                 255

ctg agc atg ctg tat gtg gat aat ggc aga gtg ctc cta gat cac cat   816
Leu Ser Met Leu Tyr Val Asp Asn Gly Arg Val Leu Leu Asp His His
            260                 265                 270

aaa gac atg atc gtg gaa gaa tgt ggg tgc ctc tgatgacatc ctggagggag 869
Lys Asp Met Ile Val Glu Glu Cys Gly Cys Leu
        275                 280

actggatttg cctgcactct ggaaggctgg gaaactcctg gaagacatga taaccatcta 929

atccagtaag gagaaacaga gaggggcaaa gttgctctgc ccaccagaac tgaagaggag 989

gggctgccca ctctgtaaat gaagggctca gtggagtctg gccaagcaca gaggctgctg 1049

tcaggaagag ggaggaagaa gcctgtgcag ggggctggct ggatgttctc tttactgaaa 1109

agacagtggc aaggaaaagc aaaaaaaaaa aaaaaaaaaa aaaaaaa               1156


<210> 2
<211> 283
<212> PRT
<213> Homo sapiens

<400> 2
Asp Val Ala Val Asp Gly Gln Asn Trp Thr Phe Ala Phe Asp Phe Ser
  1               5                  10                  15
```

```
Phe Leu Ser Gln Gln Glu Asp Leu Ala Trp Ala Glu Leu Arg Leu Gln
            20                  25                  30

Leu Ser Ser Pro Val Asp Leu Pro Thr Glu Gly Ser Leu Ala Ile Glu
         35                  40                  45

Ile Phe His Gln Pro Lys Pro Asp Thr Glu Gln Ala Ser Asp Ser Cys
     50                  55                  60

Leu Glu Arg Phe Gln Met Asp Leu Phe Thr Val Thr Leu Ser Gln Val
 65                  70                  75                  80

Thr Phe Ser Leu Gly Ser Met Val Leu Glu Val Thr Arg Pro Leu Ser
                 85                  90                  95

Lys Trp Leu Lys Arg Pro Gly Ala Leu Glu Lys Gln Met Ser Arg Val
            100                 105                 110

Ala Gly Glu Cys Trp Pro Arg Pro Pro Thr Pro Pro Ala Thr Asn Val
        115                 120                 125

Leu Leu Met Leu Tyr Ser Asn Leu Ser Gln Glu Gln Arg Gln Leu Gly
    130                 135                 140

Gly Ser Thr Leu Leu Trp Glu Ala Glu Ser Ser Trp Arg Ala Gln Glu
145                 150                 155                 160

Gly Gln Leu Ser Trp Glu Trp Gly Lys Arg His Arg Arg His His Leu
                165                 170                 175

Pro Asp Arg Ser Gln Leu Cys Arg Lys Val Lys Phe Gln Val Asp Phe
            180                 185                 190

Asn Leu Ile Gly Trp Gly Ser Trp Ile Ile Tyr Pro Lys Gln Tyr Asn
        195                 200                 205

Ala Tyr Arg Cys Glu Gly Glu Cys Pro Asn Pro Val Gly Glu Glu Phe
    210                 215                 220

His Pro Thr Asn His Ala Tyr Ile Gln Ser Leu Leu Lys Arg Tyr Gln
225                 230                 235                 240

Pro His Arg Val Pro Ser Thr Cys Cys Ala Pro Val Lys Thr Lys Pro
                245                 250                 255

Leu Ser Met Leu Tyr Val Asp Asn Gly Arg Val Leu Leu Asp His His
            260                 265                 270

Lys Asp Met Ile Val Glu Glu Cys Gly Cys Leu
        275                 280
```

<210> 3
<211> 1688
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (53)..(1150)

<220>
<221> mat_peptide
<222> (107)..(1150)

<220>
<221> sig_peptide
<222> (53)..(106)

<400> 3
gccttctcaa gggacagccc cactctgcct cttgctcctc cagggcagca cc atg cag 58
Met Gln

ccc ctg tgg ctc tgc tgg gca ctc tgg gtg ttg ccc ctg gcc agc ccc 106
Pro Leu Trp Leu Cys Trp Ala Leu Trp Val Leu Pro Leu Ala Ser Pro
-15 -10 -5 -1

ggg gcc gcc ctg acc ggg gag cag ctc ctg ggc agc ctg ctg cgg cag 154
Gly Ala Ala Leu Thr Gly Glu Gln Leu Leu Gly Ser Leu Leu Arg Gln
1 5 10 15

ctg cag ctc aaa gag gtg ccc acc ctg gac agg gcc gac atg gag gag 202
Leu Gln Leu Lys Glu Val Pro Thr Leu Asp Arg Ala Asp Met Glu Glu
20 25 30

ctg gtc atc ccc acc cac gtg agg gcc cag tac gtg gcc ctg ctg cag 250
Leu Val Ile Pro Thr His Val Arg Ala Gln Tyr Val Ala Leu Leu Gln
35 40 45

cgc agc cac ggg gac cgc tcc cgc gga aag agg ttc agc cag agc ttc 298
Arg Ser His Gly Asp Arg Ser Arg Gly Lys Arg Phe Ser Gln Ser Phe
50 55 60

cga gag gtg gcc ggc agg ttc ctg gcg ttg gag gcc agc aca cac ctg 346
Arg Glu Val Ala Gly Arg Phe Leu Ala Leu Glu Ala Ser Thr His Leu
65 70 75 80

ctg gtg ttc ggc atg gag cag cgg ctg ccg ccc aac agc gag ctg gtg 394
Leu Val Phe Gly Met Glu Gln Arg Leu Pro Pro Asn Ser Glu Leu Val
85 90 95

cag gcc gtg ctg cgg ctc ttc cag gag ccg gtc ccc aag gcc gcg ctg 442
Gln Ala Val Leu Arg Leu Phe Gln Glu Pro Val Pro Lys Ala Ala Leu
100 105 110

cac agg cac ggg cgg ctg tcc ccg cgc agc gcc cgg gcc cgg gtg acc 490
His Arg His Gly Arg Leu Ser Pro Arg Ser Ala Arg Ala Arg Val Thr
115 120 125

gtc gag tgg ctg cgc gtc cgc gac gac ggc tcc aac cgc acc tcc ctc 538
Val Glu Trp Leu Arg Val Arg Asp Asp Gly Ser Asn Arg Thr Ser Leu
130 135 140

atc gac tcc agg ctg gtg tcc gtc cac gag agc ggc tgg aag gcc ttc 586
Ile Asp Ser Arg Leu Val Ser Val His Glu Ser Gly Trp Lys Ala Phe
145 150 155 160

gac gtg acc gag gcc gtg aac ttc tgg cag cag ctg agc cgg ccc cgg 634
Asp Val Thr Glu Ala Val Asn Phe Trp Gln Gln Leu Ser Arg Pro Arg
165 170 175

cag ccg ctg ctg cta cag gtg tcg gtg cag agg gag cat ctg ggc ccg 682
Gln Pro Leu Leu Leu Gln Val Ser Val Gln Arg Glu His Leu Gly Pro

```
            180                 185                 190

ctg gcg tcc ggc gcc cac aag ctg gtc cgc ttt gcc tcg cag ggg gcg   730
Leu Ala Ser Gly Ala His Lys Leu Val Arg Phe Ala Ser Gln Gly Ala
        195                 200                 205

cca gcc ggg ctt ggg gag ccc cag ctg gag ctg cac acc ctg gac ctt   778
Pro Ala Gly Leu Gly Glu Pro Gln Leu Glu Leu His Thr Leu Asp Leu
    210                 215                 220

ggg gac tat gga gct cag ggc gac tgt gac cct gaa gca cca atg acc   826
Gly Asp Tyr Gly Ala Gln Gly Asp Cys Asp Pro Glu Ala Pro Met Thr
225                 230                 235                 240

gag ggc acc cgc tgc tgc cgc cag gag atg tac att gac ctg cag ggg   874
Glu Gly Thr Arg Cys Cys Arg Gln Glu Met Tyr Ile Asp Leu Gln Gly
                245                 250                 255

atg aag tgg gcc gag aac tgg gtg ctg gag ccc ccg ggc ttc ctg gct   922
Met Lys Trp Ala Glu Asn Trp Val Leu Glu Pro Pro Gly Phe Leu Ala
            260                 265                 270

tat gag tgt gtg ggc acc tgc cgg cag ccc ccg gag gcc ctg gcc ttc   970
Tyr Glu Cys Val Gly Thr Cys Arg Gln Pro Pro Glu Ala Leu Ala Phe
        275                 280                 285

aag tgg ccg ttt ctg ggg cct cga cag tgc atc gcc tcg gag act gac   1018
Lys Trp Pro Phe Leu Gly Pro Arg Gln Cys Ile Ala Ser Glu Thr Asp
    290                 295                 300

tcg ctg ccc atg atc gtc agc atc aag gag gga ggc agg acc agg ccc   1066
Ser Leu Pro Met Ile Val Ser Ile Lys Glu Gly Gly Arg Thr Arg Pro
305                 310                 315                 320

cag gtg gtc agc ctg ccc aac atg agg gtg cag aag tgc agc tgt gcc   1114
Gln Val Val Ser Leu Pro Asn Met Arg Val Gln Lys Cys Ser Cys Ala
                325                 330                 335

tcg gat ggt gcg ctc gtg cca agg agg ctc cag cca taggcgccta        1160
Ser Asp Gly Ala Leu Val Pro Arg Arg Leu Gln Pro
            340                 345

gtgtagccat cgagggactt gacttgtgtg tgtttctgaa gtgttcgagg gtaccaggag 1220

agctggcgat gactgaactg ctgatggaca aatgctctgt gctctctatg agccctgaat 1280

ttgcttcctc tgacaagtta cctcacctaa tttttgcttc tcaggaatga gaatctttgg 1340

ccactggaga gcccttgctc agttttctct attcttatta ttcactgcac tatattctaa 1400

gcacttacat gtggagatac tgtaacctga gggcagaaag cccaatgtgt cattgtttac 1460

ttgtcctgtc actggatctg ggctaaagtc ctccaccacc actctggacc taagacctgg 1520

ggttaagtgt gggttgtgca tccccaatcc agataataaa gactttgtaa aacatgaata 1580

aaacacattt tattctaaaa aaaaaaacgg cacgaggggg ggcccggtac ccaattcgcc 1640

ctatagtgag tcgtattaca attcactggc cgtcgtttta caacgtcg              1688
```

<210> 4

<211> 366
<212> PRT
<213> Homo sapiens

<400> 4

```
Met Gln Pro Leu Trp Leu Cys Trp Ala Leu Trp Val Leu Pro Leu Ala
           -15                 -10                  -5

Ser Pro Gly Ala Ala Leu Thr Gly Glu Gln Leu Leu Gly Ser Leu Leu
     -1   1                 5                  10

Arg Gln Leu Gln Leu Lys Glu Val Pro Thr Leu Asp Arg Ala Asp Met
 15                  20                  25                  30

Glu Glu Leu Val Ile Pro Thr His Val Arg Ala Gln Tyr Val Ala Leu
                35                  40                  45

Leu Gln Arg Ser His Gly Asp Arg Ser Arg Gly Lys Arg Phe Ser Gln
            50                  55                  60

Ser Phe Arg Glu Val Ala Gly Arg Phe Leu Ala Leu Glu Ala Ser Thr
         65                  70                  75

His Leu Leu Val Phe Gly Met Glu Gln Arg Leu Pro Pro Asn Ser Glu
     80                  85                  90

Leu Val Gln Ala Val Leu Arg Leu Phe Gln Glu Pro Val Pro Lys Ala
 95                  100                 105                 110

Ala Leu His Arg His Gly Arg Leu Ser Pro Arg Ser Ala Arg Ala Arg
                115                 120                 125

Val Thr Val Glu Trp Leu Arg Val Arg Asp Asp Gly Ser Asn Arg Thr
            130                 135                 140

Ser Leu Ile Asp Ser Arg Leu Val Ser Val His Glu Ser Gly Trp Lys
        145                 150                 155

Ala Phe Asp Val Thr Glu Ala Val Asn Phe Trp Gln Gln Leu Ser Arg
    160                 165                 170

Pro Arg Gln Pro Leu Leu Leu Gln Val Ser Val Gln Arg Glu His Leu
175                 180                 185                 190

Gly Pro Leu Ala Ser Gly Ala His Lys Leu Val Arg Phe Ala Ser Gln
                195                 200                 205

Gly Ala Pro Ala Gly Leu Gly Glu Pro Gln Leu Glu Leu His Thr Leu
            210                 215                 220

Asp Leu Gly Asp Tyr Gly Ala Gln Gly Asp Cys Asp Pro Glu Ala Pro
        225                 230                 235

Met Thr Glu Gly Thr Arg Cys Cys Arg Gln Glu Met Tyr Ile Asp Leu
    240                 245                 250

Gln Gly Met Lys Trp Ala Glu Asn Trp Val Leu Glu Pro Pro Gly Phe
255                 260                 265                 270

Leu Ala Tyr Glu Cys Val Gly Thr Cys Arg Gln Pro Pro Glu Ala Leu
                275                 280                 285
```

```
Ala Phe Lys Trp Pro Phe Leu Gly Pro Arg Gln Cys Ile Ala Ser Glu
            290                 295                 300

Thr Asp Ser Leu Pro Met Ile Val Ser Ile Lys Glu Gly Gly Arg Thr
        305                 310                 315

Arg Pro Gln Val Val Ser Leu Pro Asn Met Arg Val Gln Lys Cys Ser
    320                 325                 330

Cys Ala Ser Asp Gly Ala Leu Val Pro Arg Arg Leu Gln Pro
335                 340                 345


<210> 5
<211> 354
<212> PRT
<213> Homo sapiens

<400> 5
Met Ser Ala His Ser Leu Arg Ile Leu Leu Leu Gln Ala Cys Trp Ala
  1               5                  10                  15

Leu Leu His Pro Arg Ala Pro Thr Ala Ala Ala Leu Pro Leu Trp Thr
            20                  25                  30

Arg Gly Gln Pro Ser Ser Pro Ser Pro Leu Ala Tyr Met Leu Ser Leu
        35                  40                  45

Tyr Arg Asp Pro Leu Pro Arg Ala Asp Ile Ile Arg Ser Leu Gln Ala
    50                  55                  60

Gln Asp Val Asp Val Thr Gly Gln Asn Trp Thr Phe Thr Phe Asp Phe
65                  70                  75                  80

Ser Phe Leu Ser Gln Glu Glu Asp Leu Val Trp Ala Asp Val Arg Leu
                85                  90                  95

Gln Leu Pro Gly Pro Met Asp Ile Pro Thr Glu Gly Pro Leu Thr Ile
            100                 105                 110

Asp Ile Phe His Gln Ala Lys Gly Asp Pro Glu Arg Asp Pro Ala Asp
        115                 120                 125

Cys Leu Glu Arg Ile Trp Met Glu Thr Phe Thr Val Ile Pro Ser Gln
    130                 135                 140

Val Thr Phe Ala Ser Gly Ser Thr Val Leu Glu Val Thr Lys Pro Leu
145                 150                 155                 160

Ser Lys Trp Leu Lys Asp Pro Arg Ala Leu Glu Lys Gln Val Ser Ser
                165                 170                 175

Arg Ala Glu Lys Cys Trp His Gln Pro Tyr Thr Pro Pro Val Pro Val
            180                 185                 190

Ala Ser Thr Asn Val Leu Met Leu Tyr Ser Asn Arg Pro Gln Glu Gln
        195                 200                 205

Arg Gln Leu Gly Gly Ala Thr Leu Leu Trp Glu Ala Glu Ser Ser Trp
    210                 215                 220

Arg Ala Gln Glu Gly Gln Leu Ser Val Glu Arg Gly Gly Trp Gly Arg
```

225 230 235 240

Arg Gln Arg Arg His His Leu Pro Asp Arg Ser Gln Leu Cys Arg Arg
245 250 255

Val Lys Phe Gln Val Asp Phe Asn Leu Ile Gly Trp Gly Ser Trp Ile
260 265 270

Ile Tyr Pro Lys Gln Tyr Asn Ala Tyr Arg Cys Glu Gly Glu Cys Pro
275 280 285

Asn Pro Val Gly Glu Glu Phe His Pro Thr Asn His Ala Tyr Ile Gln
290 295 300

Ser Leu Leu Lys Arg Tyr Gln Pro His Arg Val Pro Ser Thr Cys Cys
305 310 315 320

Ala Pro Val Lys Thr Lys Pro Leu Ser Met Leu Tyr Val Asp Asn Gly
325 330 335

Arg Val Leu Leu Glu His His Lys Asp Met Ile Val Glu Glu Cys Gly
340 345 350

Cys Leu

<210> 6
<211> 368
<212> PRT
<213> Homo sapiens

<400> 6
Met Pro Phe Leu Trp Leu Cys Trp Ala Leu Trp Ala Leu Ser Leu Val
1 5 10 15

Ser Leu Arg Glu Ala Leu Thr Gly Glu Gln Ile Leu Gly Ser Leu Leu
20 25 30

Gln Gln Leu Gln Leu Asp Gln Pro Pro Val Leu Asp Lys Ala Asp Val
35 40 45

Glu Gly Met Val Ile Pro Ser His Val Arg Thr Gln Tyr Val Ala Leu
50 55 60

Leu Gln His Ser His Ala Ser Arg Ser Arg Gly Lys Arg Phe Ser Gln
65 70 75 80

Asn Leu Arg Glu Val Ala Gly Arg Phe Leu Val Ser Glu Thr Ser Thr
85 90 95

His Leu Leu Val Phe Gly Met Glu Gln Arg Leu Pro Pro Asn Ser Glu
100 105 110

Leu Val Gln Ala Val Leu Arg Leu Phe Gln Glu Pro Val Pro Arg Thr
115 120 125

Ala Leu Arg Arg Gln Lys Arg Leu Ser Pro His Ser Ala Arg Ala Arg
130 135 140

Val Thr Ile Glu Trp Leu Arg Phe Arg Asp Asp Gly Ser Asn Arg Thr
145 150 155 160

```
Ala Leu Ile Asp Ser Arg Leu Val Ser Ile His Glu Ser Gly Trp Lys
                165                 170                 175

Ala Phe Asp Val Thr Glu Ala Val Asn Phe Trp Gln Gln Leu Ser Arg
            180                 185                 190

Pro Arg Gln Pro Leu Leu Leu Gln Val Ser Val Gln Arg Glu His Leu
        195                 200                 205

Gly Pro Gly Thr Trp Ser Ser His Lys Leu Val Arg Phe Ala Ala Gln
    210                 215                 220

Gly Thr Pro Asp Gly Lys Gly Gln Gly Glu Pro Gln Leu Glu Leu His
225                 230                 235                 240

Thr Leu Asp Leu Lys Asp Tyr Gly Ala Gln Gly Asn Cys Asp Pro Glu
                245                 250                 255

Ala Pro Val Thr Glu Gly Thr Arg Cys Cys Arg Gln Glu Met Tyr Leu
            260                 265                 270

Asp Leu Gln Gly Met Lys Trp Ala Glu Asn Trp Ile Leu Glu Pro Pro
        275                 280                 285

Gly Phe Leu Thr Tyr Glu Cys Val Gly Ser Cys Leu Gln Leu Pro Glu
    290                 295                 300

Ser Leu Thr Ser Arg Trp Pro Phe Leu Gly Pro Arg Gln Cys Val Ala
305                 310                 315                 320

Ser Glu Met Thr Ser Leu Pro Met Ile Val Ser Val Lys Glu Gly Gly
                325                 330                 335

Arg Thr Arg Pro Gln Val Val Ser Leu Pro Asn Met Arg Val Gln Thr
            340                 345                 350

Cys Ser Cys Ala Ser Asp Gly Ala Leu Ile Pro Arg Arg Leu Gln Pro
        355                 360                 365
```

<210> 7
<211> 305
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (5)
<223> n equals a, t, g, or c

<220>
<221> misc_feature
<222> (28)
<223> n equals a, t, g, or c

<220>
<221> misc_feature
<222> (36)..(38)

<223> n equals a, t, g, or c

<220>
<221> misc_feature
<222> (44)
<223> n equals a, t, g, or c

<220>
<221> misc_feature
<222> (67)
<223> n equals a, t, g, or c

<220>
<221> misc_feature
<222> (101)
<223> n equals a, t, g, or c

<220>
<221> misc_feature
<222> (133)
<223> n equals a, t, g, or c

<220>
<221> misc_feature
<222> (149)
<223> n equals a, t, g, or c

<220>
<221> misc_feature
<222> (154)
<223> n equals a, t, g, or c

<220>
<221> misc_feature
<222> (195)
<223> n equals a, t, g, or c

<220>
<221> misc_feature
<222> (258)
<223> n equals a, t, g, or c

<220>
<221> misc_feature
<222> (272)
<223> n equals a, t, g, or c

<220>
<221> misc_feature
<222> (299)
<223> n equals a, t, g, or c

<400> 7
ggcanagcag ctcctgggca gcctgctngg cactcnnnta caangaggtg ccaaacctgg 60

acagggncga catggaggag ctggtcatcc ccacccacgt nagggaacca gtacgtggcc 120

ctgctgcagc gcncaacggg gaaccactnc ccgngaaana gaggttcagc cagagcttcc 180

ggcagccccc ggagnccctg gccttcaagt ggccgttttt ggggcctcga cagtncatcg 240

nctcggagac tgattcgntg cccatgatcg tncaacatca aggagggagg caggaccang 300

cccca 305

<210> 8
<211> 110
<212> DNA
<213> Homo sapiens

<400> 8
tcaagggngc agccccactc tgcctcttgn tccttccagg ggtagcacca tgcagcccct 60

gtggatctgc tgggcactct gggtgttgcc cctgggcann cccggggncn 110

<210> 9
<211> 29
<212> DNA
<213> Homo sapiens

<400> 9
cgcggatccc atcacttgcc agacagaag 29

<210> 10
<211> 42
<212> DNA
<213> Homo sapiens

<400> 10
gtacgcaagc ttgcaggcaa atccagtctc cctccaggga tg 42

<210> 11
<211> 36
<212> DNA
<213> Homo sapiens

<400> 11
caattggatc cacttgccag acagagaact caactg 36

<210> 12
<211> 39
<212> DNA
<213> Homo sapiens

<400> 12
cacttaggta ccatgtcatc agaggcaccc acattcttc 39

<210> 13
<211> 131
<212> DNA
<213> Homo sapiens

<400> 13
gccggatccg ccaccatgaa ctccttctcc acaagcgcct tcggtccagt tgccttctcc 60

ctggggctgc tcctggtgtt gcctgctgcc ttccctgccc cagtcatcac ttgccagaca 120

gaagtcaact g 131

<210> 14
<211> 36
<212> DNA
<213> Homo sapiens

<400> 14
ggctctagaa tgtcatcaga ggcacccaca ttcttc 36

<210> 15
<211> 36
<212> DNA
<213> Homo sapiens

<400> 15
gactggatcc catacttgcc agacagaagt caactg 36

<210> 16
<211> 39
<212> DNA
<213> Homo sapiens

<400> 16
cacttaggta ccatgtcatc agaggcaccc acattcttc 39

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/17211

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12P 21/00; C12N 1/21, 5/10, 15/00; C07H 21/04; C07K 14/46
US CL :435/69.1, 252.3, 325, 320.1; 536/23.4; 530/350
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 252.3, 325, 320.1; 536/23.4; 530/350

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, CAPLUS
search terms: lefty, tgf-beta, situs inversus, l r determination, l r polarity

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	MENO et al. Left-right asymmetric expression of the TGF-beta family member lefty in mouse embryos. Nature. 09 May 1996, Vol. 381, pages 151-155, see especially Figures 1 and 2.	1, 13, 18, 19, 22-27, 29, 31 --- 2, 4, 6, 8-11, 14-17, 21

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 08 NOVEMBER 1998	Date of mailing of the international search report 02 DEC 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer David S. Romeo D. Lawrence for Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)★

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/17211

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1, 2, 4, 6, 8-11, 13-19, 21-27, 29 and 31

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/17211

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1, 2, 4, 6, 8-11, 13-19, 21-27, 29 and 31, to the extent that they are drawn to Lefty nucleic acid molecules, polypeptides, and methods of making Lefty.
Group II, claim(s) 1, 2, 3, 5, 7, 12, 14, 15, 17-20, 22-28, 31, to the extent that they are drawn to Nodal nucleic acid molecules, polypeptides and methods of making Nodal.
Group III, claim(s) 30 and 32, to the extent that they are drawn to an antibody to a Lefty polypeptide.
Group IV, claim(s) 30 and 32, to the extent that they are drawn to an antibody to a Nodal polypeptide.
Group V, claim(s) 33, to the extent that it is drawn to a method of gene therapy with a polynucleotide encoding Lefty.
Group VI, claim(s) 33, to the extent that it is drawn to a method of gene therapy with a polynucleotide encoding Nodal.
Group VII, claim(s) 34, to the extent that it is drawn to a diagnostic process involving a Lefty polynucleotide.
Group VIII, claim(s) 34, to the extent that it is drawn to a diagnostic process involving a Nodal polynucleotide.
Group IX, claim(s) 35, to the extent that it is drawn to a diagnostic process involving a Lefty polypeptide.
Group X, claim(s) 35, to the extent that it is drawn to a diagnostic process involving Nodal polypeptide.
Group XI, claim(s) 36, to the extent that it is drawn to an agonist of Lefty.
Group XII, claim(s) 36, to the extent that it is drawn to an antagonist of Lefty.
Group XIII, claim(s)36, to the extent that it is drawn to an agonist of Nodal.
Group XIV, claim(s) 36, to the extent that it is drawn to an antagonist of Nodal.

The inventions listed as Groups I-II do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the Lefty and Nodal polynucleotides and polypeptides are structurally and functionally distinct compounds each of which can be made and used without the other compound; lack of unity of invention is shown because these compounds lack a common utility which is based upon a common structural feature which has been identified as the basis for that common utility.

Pursuant to 37 CFR 1.475(d), this authority considers that where multiple products and processes are claimed, the first recite product, method of making that product, and method of using that product, together with the first recited of each of the other inventions related thereto, shall constitute the main invention. Further, it considers that any subsequently recited products and/or methods constitute separate groups. Accordingly, Groups III-XIV constitute separate groups.